Means

(1)

AD-A210 746

Analysis of Aptitude, Training, and Job Performance Measures

FINAL REPORT

By

Michael P. Wagner, Ph.D.
Robert P. Dirmeyer
Barbara Means, Ph.D.
Margery K. Davidson

February 1982

Contract #MDA 903-80-C-0440
MGA-81-1080(2)-WRO-0440

Monitored Technically By:

Dr. W. S. Sellman
Office of the Assistant Secretary of Defense
(Manpower, Reserve Affairs, and Logistics)
The Pentagon
Washington, D.C.

McFann · Gray & Associates Inc.

130

SECURITY CLASSIFICATION OF T

REPORT DO

AD-A210 746

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

ECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

FINAL REPORT: Analysis of Aptitude, Training and Job Performance Measures

5. TYPE OF REPORT & PERIOD COVERED

FINAL REPORT
16 June 80 - 28 Feb. 82

6. PERFORMING ORG. REPORT NUMBER

MGA-81-1080(2)-WRO-0440

7. AUTHOR(s)

M.P. Wagner B.M. Means
R.P. Dirmeyer M.K. Davidson

8. CONTRACT OR GRANT NUMBER(s)

MDA 903-80-C-0440

9. PERFORMING ORGANIZATION NAME AND ADDRESS

McFann-Gray and Associates, Inc. (MGA)
2100 Garden Road, Suite J
Monterey, California 93940 (408/373-1111)

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

OASD(MRA&L) Room 2B269 The Pentagon
Washington, DC 20301 ATTN: Dr. W.S. Sellman

12. REPORT DATE

February 1982

13. NUMBER OF PAGES

259

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

Same, Block 11

15. SECURITY CLASS. (of this report)

UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

AFQT
ASVAB (
Attrition,
Job Performance,
Skill Qualification Tests (SQTs),
Training Performance,
Validation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This study relates AFQT and Aptitude Composite scores to measures of performance in training and on the job. The utility of current performance measures is evaluated. Alternate measures, which are available but not currently utilized, are identified and assessed. In addition, experimental performance measures were developed and tried out to determine their potential as performance measures. The relationship between experimental measures and AFQT/ASVAB was then determined. Finally, recommendations are made:

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

1) regarding the use and/or improvement of current training and job performance measures; and

2) concerning the potential of alternate experimental performance measures as criteria against which AFQT/ASVAB can be validated.

ANALYSIS OF APTITUDE, TRAINING, AND JOB PERFORMANCE MEASURES

FINAL REPORT

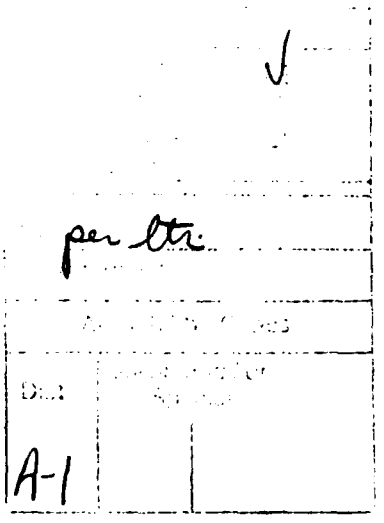
By

Michael P. Wagner, Ph.D.
Robert P. Dirmeyer
Barbara Means, Ph.D.
Margery K. Davidson

McFann-Gray & Associates, Inc.
Washington Regional Office
Arlington, Virginia

February 1982

Contract #MDA 903-80-C-0440
MGA-81-1080(2)-WRO-0440



Monitored Technically By:

Dr. W. S. Sellman
Office of the Assistant Secretary of Defense
(Manpower, Reserve Affairs, and Logistics)
The Pentagon
Washington, D.C.

CONTRACT NUMBER:	MDA 903-80-C-0440
CONTRACT EXPIRATION DATE:	28 FEBRUARY 1982
PROGRAM TITLE:	ANALYSIS OF APTITUDE, TRAINING, AND JOB PERFORMANCE MEASURES
CONTRACTOR:	McFANN-GRAY & ASSOCIATES, INC. WASHINGTON REGIONAL OFFICE 2020 NO. 14TH STREET, SUITE 409 ARLINGTON, VIRGINIA 22201
PROJECT DIRECTOR:	ROBERT P. DIRMEYER
PHONE NUMBER:	(703) 525-4773
DELIVERABLE:	Final Report 0002AF

EXECUTIVE SUMMARY

PURPOSE

One purpose of this study was to assess the utility of current performance measures in the Army (eight specialties) and Marine Corps (two specialties) for validating enlistment tests (i.e., ASVAB). In order to accomplish this task (1) current measures of performance in initial entry training courses were examined, (2) job performance measures in the Army (i.e., SQT) were evaluated, and (3) the relationship between these training and job performance measures and ASVAB scores was determined.

A second purpose of this research was to identify and/or develop alternative existing or experimental performance measures with potential utility as criteria against which enlistment tests can be validated. Five alternative measures -- three existing measures and two experimental measures -- were obtained on samples from selected Army MOSs, and were evaluated with respect to their relationship to ASVAB scores on the same samples.

RESULTS

Current Job Performance Measures

1. SQT scores are positively related to AFQT and to aptitude composite scores; that is, recruits with high AFQT/ASVAB scores also score high on the SQT.

2. The SQT is positively related to graduation from high school. High school graduates perform better on the SQT than non-high school graduates with equivalent AFQT/ASVAB scores.

3. The relationship between AFQT/ASVAB and SQT scores is about the same for whites, blacks, and Hispanics. AFQT/ASVAB scores tend to overpredict the SQT performance of blacks (and to a lesser extent Hispanics).

4. Scores on the SQT increase by an average of 20 percent from the first to the second year of fielding. This improvement suggests that there has been an increase in command emphasis on training those skills tested on the SQT, or that the tests are partially compromised in the process of field administration.

5. Existing conditions which limit the effectiveness of SQT as a criterion include:

    - o the current self-rating method of identifying performers and nonperformers during the field trials of SQT items

- o the failure to consider reliability or task complexity in determining the number of items or performance measures used to test each task
- o the insufficient training of test writing personnel
- o test compromise caused by SQT Notices and by advanced practice sessions on the Hands-On Component
- o the leniency of supervisor ratings on the Job-Site Component.

Current Training Performance Measures

1. The relationship between AFQT/ASVAB scores and final course grades is stronger for technical and administrative occupational specialties than for combat arms occupational specialties.
2. High school and non-high school graduates with the same AFQT/ASVAB scores score at about the same levels on final course grades.
3. Attrition is higher among non-high school graduates than among high school graduates.
4. Time-to-complete indices are moderately correlated with AFQT/ASVAB scores. The value of such indices, however, may depend on the presence of sufficient incentives for trainees to finish training as quickly as possible.

Alternative and Experimental Performance Measures

1. Experimental job performance ratings indicated that
   - o there was substantial disagreement between soldiers and their supervisors regarding the tasks performed by individual soldiers,
   - o soldiers perform tasks that vary substantially in degree of difficulty, and
   - o job performance ratings were variable enough to distinguish among various levels of performance.
2. In an experimental setting, peer nomination ratings during initial entry training were moderately correlated with AFQT/ASVAB scores. High school graduates were rated higher than non-high school graduates.

RECOMMENDATIONS

Recommendations are offered in light of the purposes of this study which were to (1) determine the utility of existing training and job performance measures for validating AFQT/ASVAB and (2) develop experimental alternative training and job performance measures which have potential as criteria for validating AFQT/ASVAB.

The SQT is a valuable criterion of job performance in the Army. The implementation of the SQT has apparently spurred unit training in MOS-specific tasks, resulting in the improved performance of skill level 1 soldiers. While SQT results correlate substantially with AFQT/ASVAB scores, a number of deficiencies in the SQT system were identified, which, if remedied, would enhance the value of SQTs to the Army.

- SQT Notices should contain only a sample of tasks to be tested on the subsequent operational administration of the SQT and should not specify, even for these tasks, the exact nature of the test.

- Item selection procedures on the SQT tryout should be based on pre- and post-training discrimination indices or on measures of internal consistency rather than on self-ratings (which is currently the predominant method), at least until better methods are developed (MGA is currently working on a project for the Army Training Support Center to develop more effective procedures).

- Item selection criteria should be changed; difficult items (p values less than .50) should not be automatically excluded from the test, and the criteria for item selection should include a requirement that each item significantly discriminate performers from nonperformers (rather than the current method of simply requiring equal or higher scores from performers than nonperformers).

- Empirical procedures for setting SQT cutoff scores on task tests are effective at linking test performance to performance standards only when performers and nonperformers can be accurately identified. In the absence of such accuracy, subject matter experts should assess the adequacy of task test cutoff scores.

- The practice of having at least two administrations of the HOC, with only the latter administration being operationally scored, should be changed. The first administration should be operationally scored. Subsequent administrations could be used to assess the effects of training.

- The actual scores (number of items correct, number of performance measures receiving a Go) on task tests should be retained in calculating total SQT scores.

- The JSC, rather than requiring Go/No Go judgments, should be changed to a multilevel scale (e.g., five points) with behavioral descriptions at each point.

- Training of SQT item writers should be expanded, particularly in the areas of task analysis and technical evaluation of items.

- Greater flexibility should be allowed in determining the most appropriate mix of test methods (i.e., SC, HOC, or JSC) for an MOS. Further, consideration might be given to the idea of putting more emphasis on testing specialty-specific tasks in those occupations in which (1) job

content remains fairly stable, necessitating less extensive test modifications, and (2) less specialization occurs on the job, making tests more acceptable and relevant to examinees. In general, combat arms MOSs meet these requirements to a greater extent than do combat support/combat service support specialties. The need for performance measurement in combat support/combat service support specialties might best be satisfied by developing more generic task tests for the SQT which (1) will not be acutely sensitive to changes in job content, and (2) will be relevant to examinees who specialize in their jobs.

Several recommendations are also offered with regard to training criteria.

- Serious questions have been raised concerning training criteria (i.e., final course grades), particularly with regard to the lack of validation studies. This research provides an ideal opportunity to conduct such studies. For example, the training samples used in this research could be followed to determine their success on the job. The resulting data would help to determine the predictive validity of existing training criteria as well as alternative and experimental criteria developed in the present study.

- The finding of higher attrition rates for non-high school graduates as compared to high school graduates and the lack of a relationship between attrition and AFQT/ASVAB scores suggests the need to conduct further research to isolate the correlates of attrition.

- Time to complete training could provide a suitable criterion, especially if clear incentives were established for trainees to complete courses as rapidly as possible.

TABLE OF CONTENTS

LIST OF TABLES

PAGE

PAGE

LIST OF FIGURES

PREFACE

This report was prepared for the Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics), Department of Defense, by McFann-Gray and Associates, Arlington, Virginia. It assesses the quality of current training and job performance measures, examines the relationship between AFQT/ASVAB scores and training and job performance of military personnel, and explores alternative methods and measures for improving training and job performance measurement. Field research was conducted with the support of HQ FORSCOM at Fort Bragg, North Carolina and Fort Hood, Texas. Field research was also conducted at a number of TRADOC schools and at Marine Corps training centers at Camp Pendleton and Camp Twentynine Palms. This research also received considerable support from the U.S. Army Training Support Center (SQT Management Directorate in particular) and from the Defense Manpower Data Center, without whose help the data analysis could not have been completed.

SECTION I

INTRODUCTION

The purpose of this report is to provide information on:

- o the relationship between scores on the Armed Services Vocational Aptitude Battery (ASVAB) and training and job performance of military personnel.
- o the quality of current training and job performance measures.
- o alternative methods and measures for improving training and job performance evaluation.

The work was performed by McFann, Gray and Associates, Inc. (MGA) under Department of Defense (DoD) contract #MDA 903-80-C-0440, "Analysis of Aptitude, Training and Job Performance Measures".

BACKGROUND

Selection and Classification

The Armed Services Vocational Aptitude Battery (ASVAB) is given to all applicants for enlistment. It was introduced on 1 January 1976 as the common DoD test to replace aptitude test batteries then in use by each Service.

The current version of the ASVAB (Forms 8, 9, and 10) consists of ten component subtests. The Armed Forces Qualification Test (AFQT) score and various aptitude composite scores are derived from different combinations of the component tests of the ASVAB. The AFQT scores, supplemented by scores on the aptitude composites, are used to decide whether an applicant is eligible to enlist. The scores on the aptitude composites determine eligibility to enter specific military occupations.

Like other employers competing in the labor market, the Military Services have traditionally raised or lowered entrance standards in response to the availability of applicants. When enlistment standards are reduced, it is easier to recruit enough people to meet recruitment goals. However, when standards are too low training costs may increase, and the performance of units in the field may suffer.

Pencil-and-paper tests have been used for enlistment screening since the end of World War II. During World War II, men were accepted for service so long as they had completed the fourth grade or were able to pass literacy screening tests. After service entry, the primary test instrument for job assignment purposes was the Army General Classification Test (AGCT). A test of general trainability, the AGCT was composed of questions which measured vocabulary, arithmetic reasoning, and spatial ability. It was later supplemented by special tests to measure mechanical, clerical, and other aptitude areas. The AGCT was subsequently used by the Army for enlistment screening in the late 1940s and became the model for the Armed Forces Qualification Test (AFQT).

In 1950, the AFQT was introduced, also as a measure of general trainability, to determine the eligibility of draftees and volunteers to enter any of the Services. AFQT norms, or tables converting test raw scores to percentile scores, were based upon the total officer and enlisted population serving in the military under mobilization conditions during World War II. This reference population has been the basis of comparison used by DoD to track the scores of its enlisted accessions since 1950 (OASD (MRA&L) 1980a).

In order to retard cheating and to update vocabulary and contemporary references, the AFQT was continuously revised by the introduction of new forms. Each new AFQT was calibrated back to the AGCT, so that successive AFQT scores would have a constant meaning in terms of the level of trainability associated with scores on earlier test versions. Starting in 1973 and continuing through 1975, the Services were not required to use a common AFQT, however. Each Service was permitted to develop conversion tables from its own test battery as a basis for estimating an individual's AFQT score. In 1976 DoD returned to the practice of using the ASVAB as the single enlistment test for all Services, and AFQT scores were again based on a common test (ASVAB Working Group, 1980).

AFQT scores are expressed in percentiles which are intended to show how a person's score compares to the scores achieved by the population that served in World War II. For example, if a recruit now receives a 75th percentile score, it means that his score is higher than the scores achieved by 75 percent of World War II military personnel. A recruit who now receives a score of 20 ranks higher than 20 percent of the World War II population. The percentile scores range from a low of one to a high of 99. A score at the 50th percentile is average compared to the officers and enlisted personnel under arms in World War II.

AFQT Categories

For convenience, AFQT scores are grouped into five broad categories and sometimes into finer subcategories. Those in Categories I and II are above average in trainability; those in Category III are average; those in Category IV are below average; and those in Category V are markedly below average. Under current Service policy, Category V personnel are not eligible to enlist.

Problems with the Calibration of AFQT

Shortly after the ASVAB was introduced in January 1976, the Services found that, compared to previous experience, an excessive number of new recruits were scoring in the upper two AFQT categories (sixty-fifth to ninety-ninth percentiles). An adjustment was made in the scoring system that reduced the number of recruits in these two upper categories. There was no evidence at that time which indicated that a change in the bottom half of the score range (below the fiftieth percentile) was needed. However, after reviewing several independent studies, the Department of Defense concluded that there was an error in the lower portion of the ability range as well. This error affected scores on both the Armed Forces Qualification Test and the aptitude composites derived from the ASVAB. The error in the calibration of the test inflated the scores of enlistment applicants primarily in the bottom half of the ability range (OASD (MRA&L) 1980a).

Table 1, abstracted from the report cited above, shows the magnitude of the distortion caused by the calibration error for FY 1979 enlisted accessions. For

example, while the Army [illegible] Table 3 rcent of accessions to be Category IV, the lowest scori the corrected figure was 46 percent.

Table 1. Comparision of Reported and Corrected AFQT Scores
FY 1979 Non-Prior Service Accessions

AFQT Category (Percentile Range)		Total DoD		Army	
		Reported	Corrected	Reported	Corrected
I	(93-99)	4	3	3	2
II	(65-92)	25	25	17	15
IIIA	(50-64)	32	18	22	13
IIIB	(31-49)	34	24	48	24
IV	(10-30)	5	30	9	46
V	(1-9)*	-	-	-	-
TOTAL		100%	100%	100%	100%

* Category Vs are not eligible to enlist.
NOTE: May not add due to rounding.

The Department of Defense began working on new versions of the ASVAB in 1978 as part of its continuous effort to improve the ASVAB and provide fresh norms. This new ASVAB (forms 8, 9, and 10), introduced in October 1980, corrected the calibration error in the previous test (OASD (MRA&L) 1980b).

In addition to improving DoD's capability to make historic comparisons, work is underway which will enable DoD, for the first time, to compare new enlistments with the current youth population as well as with the 1944 mobilization population. Towards this end, the new ASVAB was administered to a representative sample of the nation's young men and women. This sample was developed by the National Opinion Research Center of the University of Chicago. The results of this study will be essential for managing recruiting and for making realistic judgments on recruiting results.

Performance Testing

In the past the thrust of DoD's efforts to validate entrance tests has been to relate test scores to performance in training. This work was very useful because training is the first screen that recruits must pass. The assumption was also made that success in training was a good indicator of future job performance.

However, the discovery of the calibration error in the scoring of ASVAB 5, 6, and 7 raised concerns in DoD and Congress about the ability of those people whose test scores were inflated to complete training successfully and, no less significant, to perform satisfactorily on the job.

Ideally, enlistment tests should be validated against actual job performance. Enlistment standards should be based on the probability of successful job performance and the costs and benefits associated with higher or lower cut-off scores. Fully satisfactory job performance measures are not readily available. In a memorandum dated 11 September 1980, subject "Enlistment Standards", the Assistant Secretary of Defense for Manpower, Reserve Affairs, and Logistics asked the Services to begin to develop improved methods of measuring job performance. This will be a long term research effort. Future research may involve developing new job performance tests, refining existing measures of performance, or developing composites of existing measures. In the interim, decisions still have to be made on enlistment standards. These decisions can have a profound effect on manpower costs, the sustainability of the volunteer force, and the effectiveness of military units.

In February 1980, the Office of the Assistant Secretary of Defense for Manpower, Reserve Affairs, and Logistics asked MGA to undertake a study analyzing existing data on the performance of those soldiers who would have failed the AFQT and been denied enlistment if the test had been calibrated correctly. This study, conducted by I. M. Greenberg (1980), provided the context for the current research presented in this report. (The Greenberg study is contained in a report by OASD (MRA&L) 1980b.) The study concentrated on the Army because that Service enlisted the highest proportion of low-scoring recruits.

The Greenberg study was not designed to validate the ASVAB or the AFQT portion of the ASVAB. Instead, the purpose was to provide information which would assist DoD in deciding how to set enlistment standards after the test calibration error was corrected by the introduction of the new ASVAB in October 1980. If the performance data showed that most of the people who were inadvertently enlisted (potentially ineligibles*) performed satisfactorily, it would make sense to continue accepting some of these PIs to meet recruiting goals. On the other hand, if they performed poorly, it would be advisable to establish enlistment standards which exclude this group of low-scoring applicants and provide the resources to recruit higher scoring individuals.

---

*The potentially ineligibles (PIs) are soldiers who would have been denied enlistment if correct ASVAB scoring tables had been in effect during the period 1 January 1976 to 30 September 1980. They would have failed the AFQT and would have been barred from enlistment under the Army enlistment standards then in effect. For FY 1979, 27 percent of Army non-prior service accessions would have failed a correctly calibrated AFQT.

The Greenberg study used several measures which are reasonable surrogates for job performance: training attrition, total scores on Skill Qualification Tests, first-term attrition, reenlistment eligibility, first-term retention, and promotion. The report discusses the limitations of these measures. To some extent these limitations mirror the imperfections inherent in judging and managing people. The performance measures used in the Greenberg report reflect the real world of imperfect personnel decisions.

Training Attrition

The Greenberg study analyzed attrition in 34 different entry-level technical training courses. These courses included about two-thirds of all males who received entry-level training in the Army in FY 1979.

The 34 courses provide training in a wide variety of Military Occupational Specialties (MOSs) including combat skills, equipment maintenance, communications, supply, and administration. The Greenberg study concentrated on relatively simple occupations since the purpose was to study the performance of soldiers with low scores on the AFQT. However, each of the 34 courses contained students with higher than average scores on AFQT and aptitude composites.

The overall FY1979 attrition rate in the 34 skill training courses analyzed in this study was seven percent with a nine percent failure rate for non-high school graduates. There was little variation in the attrition rate by AFQT score. The PIs had an attrition rate of nine percent. The attrition rate was low because most of the courses in the sample were not academically demanding, and the entering students had been prescreened by having completed basic training and by meeting minimum aptitude score requirements. Some of the 34 courses were of moderate difficulty (i.e., required an aptitude score of 100). In these courses the PIs had a failure rate of twelve percent.

Skill Qualification Tests (SQTs)

SQTs are performance-oriented tests administered to enlisted personnel assigned to Army units. These tests are used by the Army as a diagnostic tool to identify training deficiencies and needs. The test results are also a factor in the point system for promotion to grade E-5 and above. The SQT has three components: a Skill Component comprised of written multiple-choice questions, a Hands-On Component assessing actual performance of tasks, and a Job-Site Component in which supervisors observe the performance of tasks on the job.

The analysis of SQT results presented in the Greenberg report was based on total SQT scores in nine occupations for soldiers who entered the service in FY1977. Component scores were not available. Soldiers with higher AFQT scores and higher scores on the aptitude composites tended to score higher on the SQTs. Completion of high school had little influence on SQT scores, probably because most of the poor-performing non-high school graduates are separated from the service before they are ready to take SQTs. For the nine MOSs sampled, 67 percent of the soldiers passed their SQT. The pass rate for PIs was 58 percent.

First-Term Attrition and Reenlistment

The sample used for the analysis presented in the Greenberg report was about 98,000 males who enlisted in FY1977 for a three-year term of service.

Thirty-nine percent of the cohort failed to complete their three years of service. The first-term attrition rate for non-high school graduates (52 percent) was nearly twice as great as the rate for graduates (27 percent). The variation by AFQT category within each educational level was slight. These results suggest that attrition is more closely related to personal characteristics than to aptitude. The PIs suffered a relatively high first-term attrition rate of 48 percent because most of the PIs in the sample (76 percent) were non-high school graduates.

Those soldiers who complete their initial term of service are considered for reenlistment. Unit commanders are responsible for deciding which soldiers are eligible to reenlist. Army regulations provide criteria and guidelines with respect to age, citizenship, medical fitness, moral suitability, trainability, and competence. There are procedures for waiving some of the requirements or providing more time to meet them. The guiding principle is that reenlistment is a privilege to be reserved only for those whose performance, conduct, attitude, and potential for advancement are in consonance with the quality standards of the Army.

Seventy-four percent of those who completed their first-term of service were eligible to reenlist. The reenlistment eligibility rate for non-high school graduates was slightly lower (70 percent) than the eligibility rate for high school graduates (76 percent). The rate did not vary much by AFQT within each educational level. The eligibility rate for the PIs was 72 percent.

Fifty-five percent of those who completed their first term and were eligible to reenlist decided to stay in the Army. The propensity to reenlist was highest for non-high school graduates (59 percent) and for PIs within each educational group (62 percent). Attrition, reenlistment eligibility, and reenlistment rate data are summarized in Figure 1 on the following page.

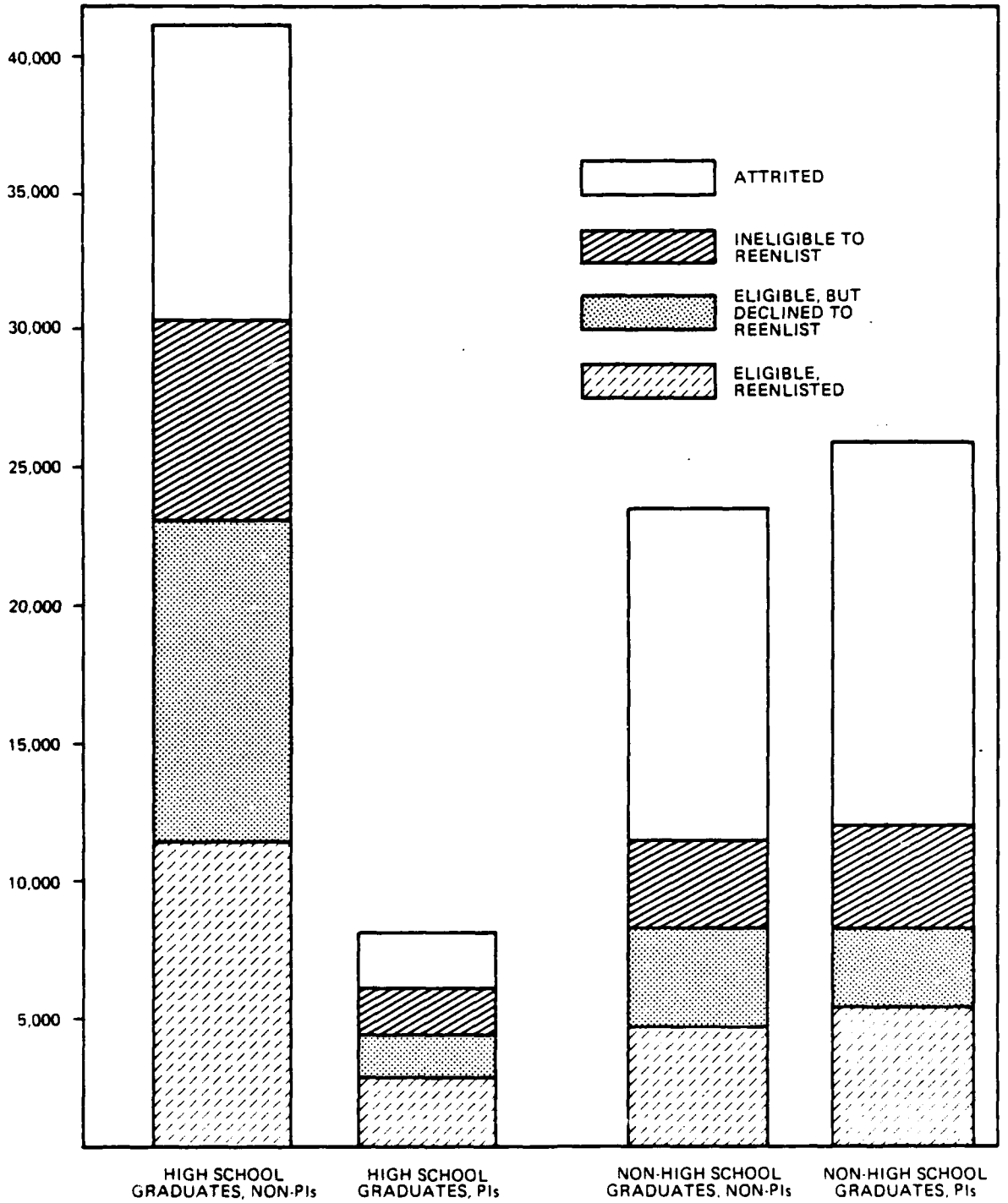
Promotion

High school graduates and those with higher AFQT scores were much more likely to be promoted to grade E-5 during their first enlistment and during the first year after they reenlisted, according to the Greenberg study. However, soldiers in all AFQT categories and educational levels had a high rate of success in achieving at least grade E-4, if they completed their first term. Seventy-six percent of the PIs who completed their first term of service and separated were E-4 or higher. The comparable rate for Category IIIA personnel was 84 percent.

As stated previously, the Greenberg study employed existing and available information. Imperfections and limitations of performance measures employed were discussed to include suggested courses of action to assist DoD in further determining the utility of training and job performance measures as criteria for validating the ASVAB and for evaluating performance of individuals.

For example, it was suggested that further research be conducted to explain why AFQT scores predict SQT performance but not entry-level training course

Figure 1
Attrition, Reenlistment Eligibility, and Reenlistment Rates for 1977 Cohort as a Function of Level of Education and Potential Ineligibility



completion for the same MOS. Some hypotheses are:

- Completion of entry-level training may not be an adequate measure of individual differences in training performance (especially in self-paced training courses). Although all graduates passed the Go/No Go tests in training, those with low AFQT/ASVAB scores may be less proficient than others or may have required more time in training.

- Soldiers with low AFQT/ASVAB scores may forget more rapidly what they learned in their entry-level training courses. They may need more refresher training in units than they are currently receiving.

- To perform well on a SQT, a soldier must master new tasks which are not taught in the entry training course preceding assignment to a unit. Deficiencies in unit training may be more of an impediment for those who score low on the AFQT/ASVAB than for those who score higher on the enlistment tests (cf. Resnick & Glaser, 1976).

Further research is also needed on the utility of SQT scores as a measure of job performance. According to Army doctrine, SQTs are to be used in diagnosing training deficiencies but not for measuring individual job performance. Two research efforts were suggested:

- Compare performance on the Skill Component (written test) of the SQT with performance on the Hands-On and Job-Site Components for those who score low on the AFQT and aptitude composites.

- Determine the relationship between the SQT scores of individual soldiers and unit effectiveness.

PROBLEM

Decisions on cut-off scores for enlistment and assignment should be based on information which relate ASVAB test scores with measures of performance in training and on the job. The Greenberg findings suggest that soldiers with ASVAB scores slightly below the current enlistment standard (the PIs) perform comparably to other soldiers with the same educational status. However, questions have been raised concerning the validity of performance measures available for that study (e.g., training attrition and Skill Qualification Test scores). More research is needed on the usefulness of current performance measures before making changes in enlistment standards which would result in accepting applicants now barred from enlistment or in rejecting applicants now qualified for enlistment. There is also a need to begin a long-range effort to validate the ASVAB against performance in training and on the job. The success of such a test validation effort will be enhanced by the identification and development of sound performance measures.

APPROACH

The overall goal of this research is to assist DoD and the Services in refining enlistment standards so that they are more accurately based on expected

performance. Basic to achieving this goal is the availability of valid and reliable measures of training and job performance that are administered consistently and fairly to all individuals.

To address these issues, the following actions were taken: first, an assessment was made of current training and job performance measures; second, alternate measures available but not employed were identified and assessed; third, alternate measures were developed and tried out to determine their potential utility as performance measures; and fourth, the relationship or correlation between the AFQT/ASVAB and each of these measures was determined.

Assessment of Training Performance Measures

Table 2 lists the ten different occupational specialties (eight Army MOSs and two Marine Corps specialties) selected for study. Included in the table are the

Table 2. Specialties Selected for Study

Specialty	Job Title	Aptitude Composites	Minimum* Aptitude Score	Service
0311	Infantryman	CO	85	Marine Corps
11B	Infantryman	CO	85	Army
11C	Indirect Fire Infantryman	CO	85	Army
19E	Armor Crewman	CO	85	Army
05C	Radio Teletypewriter Operator	SC	95	Army
31M	Multichannel Communications Operator	EL	95	Army
2841	Ground Radio Repair	EL	100	Marine Corps
		GT	110	
67N	Utility Helicopter Repairer	MM	100	Army
73C	Finance Specialist	CL	90	Army
75B	Personnel Administration Specialist	CL	95	Army

* These aptitude scores are standard scores. For the population of non-prior service Army applicants during FY 1981 the mean performance on the aptitude composites ranged from about 84 to 87 with standard deviations ranging from 17 to 19.

ASVAB aptitude composite and minimum score required for the entry-level course in each specialty. The components of the aptitude composites required for the various specialties are presented in Appendix A. These specialties were selected for study because: (a) they represent a variety of military jobs (combat arms, equipment operation, maintenance, and administration); (b) they include many individuals who have low AFQT and ASVAB aptitude composite scores; and (c) they represent jobs across the Military Services.

Arrangements were made with the Army and Marine Corps for visits to appropriate schools* to permit data collection, observation of testing, and interviewing of key personnel responsible for test development, administration, and management of training. At each installation, all students in our sample were identified and tracked until they either graduated or were dropped from the course. Course progress records were obtained for each student. This procedure was followed until information was obtained for at least 200 entering students.

Interviews were conducted and observations made to collect information on test administration and development procedures, including test content selection, item development, and test validation. Where available, written materials were collected for later analyses. Data were obtained for performance measures available but not presently employed for student evaluation. Also, at some installations, students in our sample were administered experimental performance measures. These measures, existing and experimental, are described in Section V of this report.

For each student in our sample, AFQT/ASVAB information was obtained from the Defense Manpower Data Center (DMDC).

Assessment of Job Performance Measures

The initial plan called for the examination of job performance in relation to training performance for the 10 specialties studied in the assessment of training performance measures. This was not possible for two reasons. First, the Marine Corps at present has not developed a formal job performance system that would measure individual performance in each military occupation. Consequently, the Infantryman (0311) and Ground Radio Repair (2841) specialties could not be included in the job performance assessment. However, the Marine Corps is currently conducting a study to determine the feasibility of establishing enlistment standards and assignment criteria (school/job prerequisites) based on expected job performance. In order to conduct this feasibility research, job performance measures had to be developed for several occupational specialties. If the study establishes that the concept is feasible, job performance measures will eventually be developed for all Marine Corps occupational specialties.

Second, except for special externally imposed studies, the Army schools do not retain training performance data on individuals. Thus, no training performance data were obtainable for first-term soldiers who were serving in units at the time of the study.

* See Appendix B for schools and schedules of visit.

The approach employed was to obtain SQT score information* for the eight Army MOSs studied in training. The sample included all soldiers in these MOSs who were first-term enlistees, who took ASVAB Forms 6 or 7, and had been administered an SQT during FY 1977-81. These records were matched with ASVAB accession records provided by DMDC. These data served as the basis for relating AFQT/ASVAB to job performance (SQT) for the eight Army MOSs.

Site visits were made to the proponent agencies responsible for development and maintenance of the relevant SQTs. Key individuals were interviewed, relevant SQT materials were examined, and selected materials were collected for later analysis. As part of this information gathering process, discussions were held with the Deputy Commander for Skill Qualification Test Management and his staff at Ft. Eustis, Virginia.

In addition, visits were made to two U.S. Army Forces Command (FORSCOM) installations (Ft. Bragg, North Carolina and Ft. Hood, Texas) to observe SQT testing and to interview individuals involved with the SQT program concerning their views on job performance testing. Local policy on management and administration of the SQTs was obtained as part of the query.

Further, at both FORSCOM installations, experimental job performance measures were administered to job incumbents in three MOSs (Infantryman, Multichannel Communications Operator, Personnel Administration Specialist). Data were obtained on a total of 120 soldiers for the three MOSs combined. These data were employed to determine the relationship of the experimental measures to AFQT/ASVAB pre-entry scores.

ORGANIZATION OF REPORT

This report examines the performance measurement of regular Army and Marine Corps first-term personnel. The remainder of the report is presented in the following format:

Section II:	Relationship of AFQT/ASVAB to Job Performance Measures
Section III:	Relationship of AFQT/ASVAB to Training Performance Measures
Section IV:	Assessment of Job Performance Measures
Section V:	Assessment of Training Performance Measures
Section VI:	Assessment of Alternative Job Performance Measures
Section VII:	Summary and Recommendations

* The Army Training Support Center, Ft. Eustis, Va., maintains a master file of SQT scores including both total and component scores.

SECTION II

RELATIONSHIP OF AFQT/ASVAB SCORES TO JOB PERFORMANCE MEASURES

This section of the report relates scores on the Skill Qualification Tests to scores on the AFQT and to scores on the ASVAB aptitude composites.

SKILL QUALIFICATION TESTS (SQTs)

SQTs are administered to enlisted Army personnel. The SQT evaluates a soldier's ability to perform about 25-35 critical tasks from his Soldier's Manual. The tasks are specific to the Military Occupational Specialty (MOS) in which the soldier is serving. The SQTs are "performance oriented", which means that they focus on the skills and knowledges needed to perform tasks in a realistic job environment.

SQTs are used by the Army as diagnostic tools to identify deficiencies in the training system and the need for remedial training of individual soldiers. SQT scores are also used in the enlisted promotion system for promotion to E-5 and above. A soldier can earn a maximum of 150 points toward promotion for SQT performance in a rating system which provides for a maximum of 1000 promotion points.

There are normally three components in an SQT:

a. The Skill Component (SC) is a multiple-choice test which evaluates a soldier's ability to apply performance-relevant knowledge. The SC is a paper-and-pencil test, supplemented in some instances by audiovisual material. Soldiers record answers on a marksense answer sheet. This part of the SQT is often called the "written test."

b. The Hands-On Component (HOC) requires soldiers to perform job-relevant tasks under highly standardized conditions. The HOC usually calls for a formal test site, trained scorers, and actual equipment or simulators.

c. The Job-Site Component (JSC) is also based on hands-on performance but is administered by scoring soldiers' performance as they work on the job. The supervisor scores the soldier.

Training and Doctrine Command (TRADOC) Regulation 351-2* recommends that about one-third of the tasks on the SQT should be measured by the Job-Site Component. The mix between the Skill Component and the Hands-On Component varies by MOS and skill level. In general, the use of the Hands-On Component is large for combat MOSs, moderate for combat support MOSs, and low for combat service support MOSs. The specific guidelines for component mix are reproduced in Appendix C of this report.

*Skill Qualification Tests (SQTs), Policy and Procedures, dated 21 April 1980.

Each critical task is measured by an individual task test in one of the three SQT components. Individual task tests may consist of between two and 20 questions (Skill Component) or performances (Hands-On and Job-Site Components). A specified number of these questions/performances must be successfully answered/accomplished in order for an individual to receive a Go on a task. The percentage of tasks for which a soldier achieves a Go constitutes his/her score for the SQT. For example, a soldier who receives a score of 70 percent on an SQT has received a Go on 70 percent of the individual task tests administered to him/her. The number of task tests administered may be less than the total number of task tests prescribed for an SQT. Task tests which are not administered because of nonavailability of equipment or other reasons are scored as "not administered" and are not included in calculating the total test score.

A score of 60 percent has been set arbitrarily by the Army as a passing score on an SQT. The 60 percent score signifies that a soldier is "minimally" proficient in the specific skill level of an MOS.

Sample

The data presented in the following tables are based on all soldiers in eight Army MOSs who:

- took a skill level 1 SQT. Skill level denotes the level of qualification within a total MOS. Skill level 1 is the lowest of five designated levels and corresponds to pay grades E-3 and E-4 (it may, in some cases, include soldiers with pay grades E-1 or E-2).
- took ASVAB Form 6 or 7 upon enlistment. These forms of the ASVAB were operational from 1 January 1976 through 30 September 1980.
- completed all components of one version of an SQT. The SQT versions are annual, beginning with 1977. SQTs were not, however, fielded in all MOSs by 1977. The period for administering an SQT is roughly equivalent to the corresponding calendar year.

The sample is composed of 41,146 soldiers. The distribution of soldiers across AFQT categories is displayed for each MOS in Table 3. Furthermore, 61 percent of the soldiers in this sample (summed across all eight Army MOSs) were high school graduates and 39 percent were non-graduates. The AFQT distribution of this sample is shown in Figure 2.

SQT SCORES RELATED TO AFQT

Figures 3 through 10 (in Appendix D) show, for each MOS in the study, how soldiers with different AFQT scores performed on the three parts of the SQT: the Job-Site Component (JSC), the Hands-On Component (HOC), and the Skill Component (SC). Each point on the graphs indicates the percentage of tasks which soldiers passed. When insufficient data were available to reliably determine a point (i.e., based on fewer than 10 soldiers), that point was omitted from the figure. Figures 11

Table 3

Distribution of AFQT Scores for SQT Sample
in Eight Army MOSs

MOS	n	I	II	IIIA	IIIB	IVA	IVB
11B	24,665	2	15	12	21	23	28
11C	5,806	2	12	10	20	25	31
19E	4,142	1	13	12	21	26	28
05C	1,737	2	14	14	27	27	16
31M	2,291	0	8	10	20	25	37
67N	1,394	3	26	18	21	16	16
73C	634	3	32	23	27	12	4
75B	477	1	15	14	34	20	16

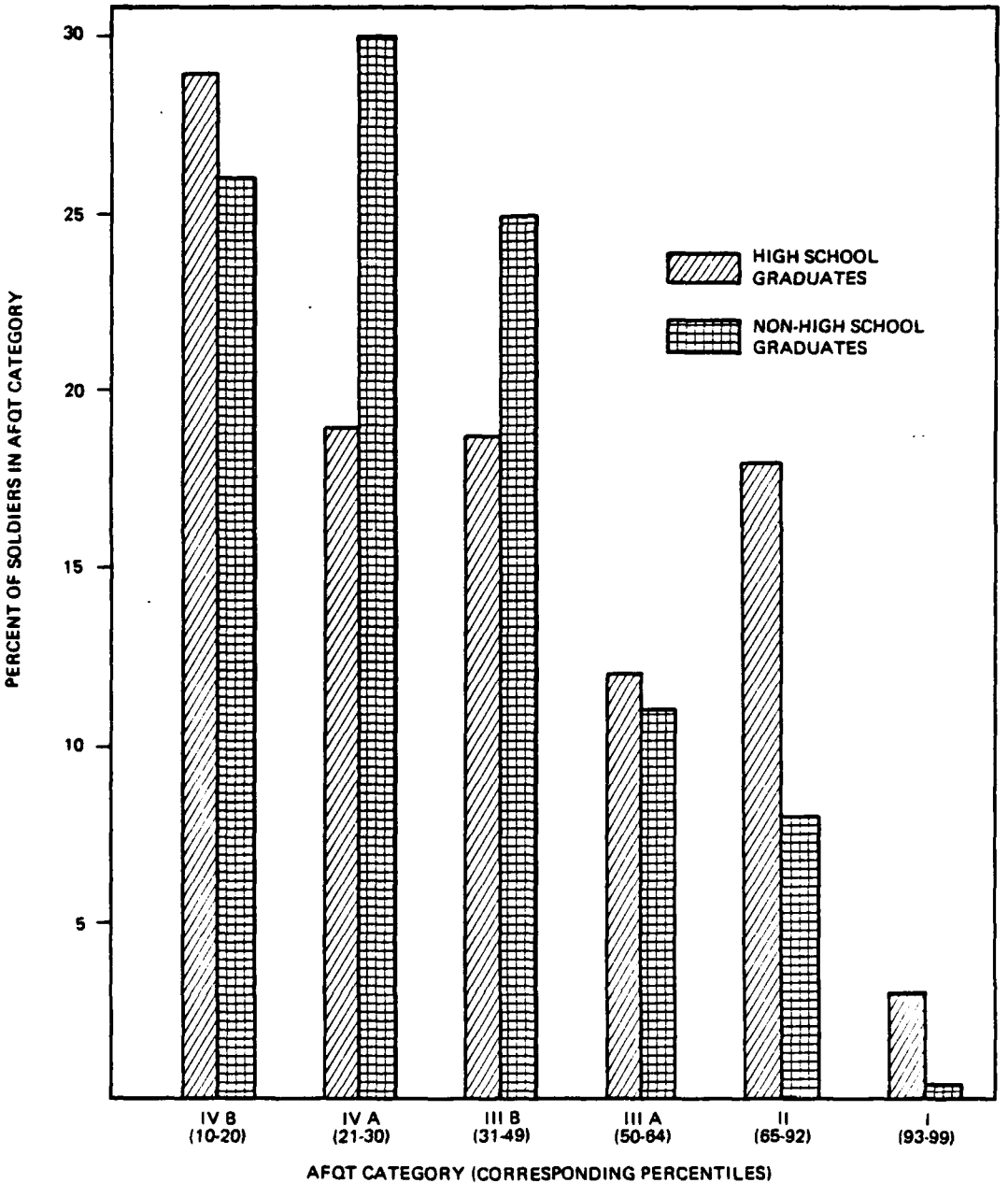
through 18 (in Appendix D) show, for each MOS in the study, separately for high school and non-high school graduates, the percent of soldiers who passed the SQT as a function of their aptitude composite scores. The correlations between SQT and AFQT scores are presented later in this section in Table 5. Once again, when a point was considered unreliable, it was omitted from the figure.

Findings

Several clear trends are evident in the data displayed in Figures 3 through 18 in Appendix D.

- Scores on the JSC were extraordinarily high across all MOSs, averaging about 98 percent. These scores did not vary as a function of AFQT scores.

Figure 2
AFQT Distribution of Sample Used in Analysis of SQT Performance as a Function of Level of Education



- Scores on the HOC were quite high for most MOSs, averaging about 85 percent, and differences between the highest and lowest scorers on AFQT averaged about 10 percent on the HOC. However, in MOS 75B (Personnel Administration Specialist), where the HOC is made up of written performance tasks (e.g., typing military correspondence), performance varied with AFQT score in a manner similar to the SC (see Figure 10).

- Scores on the SC were quite low for most MOSs, averaging less than 50 percent. These scores typically ranged from about 70 percent or higher for those soldiers who scored high on the AFQT to 40 percent or lower for soldiers who scored low on the AFQT.

- The overall pass rate on SQTs was low, averaging about 65 percent. This pass rate varied across AFQT score levels from about 80 to as low as 40 percent. As would be expected in light of the markedly low scores for the SC, the pass rate for the SQT as a whole varied negatively with the number of SC tasks included on it. For MOSs like 11B (Infantryman), with a small SC, the SQT pass rate was high, about 80 percent. Alternatively, for an MOS like 75B (Personnel Administration Specialist), with a larger SC, the pass rate was less than 50 percent.

- The pass rate for high school graduates was somewhat higher than that for non-high school graduates. For example, for 19E (Armor Crewman), 69 percent of high school graduates and 63 percent of non-high school graduates passed the SQT. However, when only soldiers who scored poorly on the AFQT were considered (i.e., category IVs), high school graduates (56 percent pass) and non-high school graduates (54 percent pass) differed only slightly.

SQT SCORES RELATED TO APTITUDE COMPOSITES

Figures 19 through 26 (in Appendix D) show, for each MOS, how soldiers with varying aptitude composite scores performed on the three components of the SQT. Figures 27 through 34 (in Appendix D) show, for each MOS, separately for high school and non-high school graduates, the percent of soldiers who passed the SQT as a function of aptitude composite scores. The correlations between SQT and aptitude composite scores are presented in Table 6 later in this section.

Findings

It is evident that the relationship of aptitude composite scores to SQT scores parallels in every respect the relationship between AFQT scores and SQT scores.

- Scores on the JSC did not vary as a function of aptitude composite scores.

- Scores on the HOC varied only about 10 percent as a function of aptitude composite scores.

- Scores on the SC ranged from about 70 percent or higher to 40 percent or lower as a function of aptitude composite scores.

- The pass rate on SQTs varied as a function of aptitude composite scores from a high of about 80 percent for soldiers with aptitude scores above 120 to as low as 40 percent for the group with aptitude scores below 80.

- Once again, while high school graduates passed the SQT at a somewhat higher rate than non-high school graduates, these differences were less marked when only low-scoring soldiers were considered. Thus, for the Armor Crewman, 69 percent of high school graduates and 63 percent of non-high school graduates passed the SQT. However, when only soldiers who scored poorly on the aptitude composites (i.e., less than 90) were considered, the pass rates for high school and non-high school graduates were 54 and 53 percent, respectively.

SQT SCORES RELATED TO YEAR ADMINISTERED

This study includes data on SQTs administered during the period 1977 through 1981. The year of initial fielding of an SQT varied across the MOSs examined in this study. Some MOSs introduced SQTs as early as 1977 (e.g., Infantryman) while others did not fully implement their SQTs until 1979 or 1980 (e.g., Personnel Administration Specialist).

In Table 4 the pass rates on SQTs are displayed for the eight MOSs in this study for the first and second year in which SQTs were fielded. This table only includes data on SQTs which were considered valid by the Army.

Findings

Several trends are evident in the data:

- Pass rates were low during the first administration of an SQT in an MOS, ranging from 12 to 67 percent. These findings have variously been attributed to either invalid, inappropriate tests or to poorly trained soldiers.

- Pass rates for the second year showed significant improvements over the first-year administration, ranging from 29 to 86 percent. This trend suggests that SQT may be having its intended effect; that is, to stimulate training. According to this view, as soldiers become more familiar with the SQT program, their preparation for subsequent tests improves. Alternatively, these data may indicate that tests are being compromised.

SQT SCORES RELATED TO AFQT/APTITUDE COMPOSITES FOR DIFFERENT RACIAL/ETHNIC GROUPS

The soldier samples in the eight Army MOSs for which SQT data were obtained were comprised of 53 percent whites, 35 percent blacks, 9 percent Hispanics, and 3

Table 4

SQT Performance in Eight Army MOSs
as a Function of Year of Administration

MOS	Job Title	First Year Fielded		Second Year Fielded	
		n	Pass Rate (Percent)	n	Pass Rate (Percent)
11B	Infantryman	2070	67	10295	83
11C	Indirect Fire Infantryman	765	29	2222	69
19E	Armor Crewman	3776	65	366*	86
05C	Radio Teletype-writer Operator	880	34	794	57
31M	Multichannel Communications Operator	1278	28	1013	55
67N	Utility Helicopter Repairer	873	12	521	29
73C	Finance Specialist	591	30	43*	35
75B	Personnel Administration Specialist	426	43	51*	63

* 1981 version of SQT. Data collection is not complete for this test version.

percent spread among other racial/ethnic groups. For the purposes of the following presentation of data, whites will compose one group, blacks will compose a second group, and Hispanics will compose a third group.

Figures 35 through 42 (in Appendix D) show, for each MOS, the SQT performance of the three racial/ethnic groups as a function of AFQT scores. (No Hispanic group appears in the figures for several MOSs which did not contain adequate numbers of soldiers from Hispanic backgrounds.) Figures 43 through 50 present comparable data as a function of aptitude composite scores. Tables 7 and 8 later in this section contain correlations between AFQT and aptitude composites and the Skill and Hands-On Components of the SQT for the three racial/ethnic groups. (Uncorrected correlation coefficients can be found in Tables 26 and 27 in Appendix E.)

Findings

It is clear that the relationship of aptitude composite scores to SQT scores for the racial/ethnic groups displayed in the figures parallels closely the relationship between AFQT scores and SQT scores for the same racial/ethnic groups.

- For the white, black, and Hispanic groups, SQT performance increased as a function of increasing aptitude composite scores. These increases in SQT pass rates across the range of aptitude composite scores ranged from 25 to 40 percent and did not differ for the three racial/ethnic groups. Thus, aptitude composite scores appear to be equally valid as predictors of SQT performance for the three ethnic groups.

- Aptitude composite scores overpredicted pass rates on SQT for the black group by between five and ten percent and for the Hispanic group, to a lesser degree. Thus, for example, when soldiers in the white and black groups with the same aptitude composite scores are compared, the pass rate on the SQT is from five to ten percent higher for the white group than for the black group. In other words, aptitude test scores for black and Hispanic groups predict higher job performance than is actually achieved.

CORRELATIONS BETWEEN AFQT AND SQT PERFORMANCE

The relationship between AFQT scores and SQT performance for the MOSs in this study is displayed in Table 5. SQT performance is indexed both by scores on each part of the SQT (Job-Site, Hands-On, and Skill Components) and by the total SQT score.

The correlation between AFQT and SQT scores of soldiers in an MOS is an estimate of the potency of AFQT as a predictor of success in that MOS. The accuracy of a correlation coefficient depends on several assumptions, including whether the sample (e.g., the soldiers in an MOS) is representative of the pool of applicants for military service from which all soldiers are drawn. If a sample contains soldiers with less variation in their AFQT scores than is found in the entire pool of applicants, any correlation based on scores of soldiers in this sample will underestimate the true relationship between AFQT and SQT scores. The range of AFQT scores for the data presented in Table 5 is restricted in at least three ways.

- o A minimum score on AFQT is required for selection into the military, thereby resulting in the elimination of low-scoring applicants from an MOS.

- o Each MOS also requires minimum scores on aptitude composites to qualify. These standards vary across MOSs. Since aptitude composites are highly correlated with AFQT scores (typically around .70) these standards have the effect of restricting the range of AFQT scores in an MOS. The extent of this restriction of range will vary across MOSs.

- o A significant amount of attrition takes place in an MOS prior to initial SQT testing. If this attrition is primarily among lower scoring individuals, it will act to reduce the range of AFQT scores in an MOS, especially in MOSs where a strong relationship exists between AFQT and SQT scores.

The correlations presented in Table 5 are corrected for restriction of range resulting from the factors listed above (Thorndike, 1949).* These correlation coefficients reflect the relationship between AFQT and SQT scores expected if soldiers were randomly selected for an MOS from among all Service applicants, and there was no attrition prior to administration of the SQT.

Findings

There are several trends evident in the data.

- o Correlations between AFQT and the Job-Site Component scores are very low, .11 or less. It is worth noting that the scores on the Job-Site Component were restricted in range (SD of about 10) relative to the scores on Hands-On (SD of about 20) and Skill Components (SD of about 20). Therefore, the resulting correlation coefficients probably underestimate the true relationship between AFQT and Job-Site Component scores.

- o Correlations between AFQT and Hands-On Component scores are generally low, ranging from .09 to .19 except for an MOS (i.e., 75B) in which the HOC is a written performance test (e.g., typing military correspondence), resulting in a higher correlation (.28).

- o Correlations between AFQT and the Skill Component are high, ranging from .46 to .55.

- o Correlations between AFQT and total SQT scores are also high, ranging from .39 to .52. However, these correlations are no higher than correlations between AFQT and Skill Component scores. The value of correlations of this magnitude is discussed in the following Section: Correlations Between Aptitude Composites and SQT Performance.

---

*Uncorrected correlation coefficients can be found in Tables 24 and 25 of Appendix E.

Table 5

Correlations* of AFQT Scores With SQT Performance for Eight Army MOSs

MOS	Job Title	n	Percent of Tasks Go: Job-Site Component	Hands-On Component	Skill Component	Total SQT
11B	Infantryman	24665	.03	.18	.52	.47
11C	Indirect Fire Infantryman	5806	.03	.19	.55	.44
19E	Armor Crewman	4142	.05	.14	.52	.47
05C	Radio Teletype-writer Operator	1737	.04	.11	.48	.47
31M	Multichannel Communications Operator	2291	.00	.09	.48	.39
67N	Utility Helicopter Repairer	1394	.04	.17	.46	.45
73C	Finance Specialist	634	.11	--**	.52	.50
75B	Personnel Administration Specialist	477	.02	.28	.51	.52

* Corrected for restrictions of range.
** Too few observations.

CORRELATIONS BETWEEN APTITUDE COMPOSITES AND SQT PERFORMANCE

Aptitude composite scores of the ASVAB are used to predict the MOSs in which an enlistee has a high probability of success. Consequently, aptitude composite scores should be correlated with job performance. In Table 6, correlations between aptitude composite scores and measures of SQT performance are displayed. These correlations are corrected for restriction of range* caused by the factors listed in the previous section of this report.

Findings

These data closely resemble the correlations between AFQT and SQT performance.

- Correlations between aptitude composite scores and Job-Site Component scores are relatively low, .20 or less. Once again, it should be noted that Job-Site Component scores are restricted in range relative to Hands-On and Skill Component scores.

- Correlations between aptitude composite and Hands-On Component scores are generally moderate, ranging from .14 to .31. The exceptions to this are MOSs in which the HOC consists of written performance tasks (i.e., 75B), resulting in a higher correlation (.34).

- Correlations between aptitude composite and Skill Component scores are also high, ranging from .49 to .62.

- Correlations between aptitude composite and total SQT scores are high, ranging from .49 to .62. These correlations are no higher than those between aptitude composites and Skill Component scores.

A correlation coefficent between a predictor (e.g., aptitude composite) and a criterion indicates the extent to which selection based on the predictor will benefit criterion performance. When the use of a selection device results in improved criterion performance, the selection device can be said to have utility; that is, the selection device allows an institution to better utilize its resources (e.g., better soldier performance on the job). Thus, as the correlation between aptitude composite scores and a criterion of job performance increases, so too will the potential utility of the predictor for benefiting criterion performance. The actual utility realized from the use of a selection device with a specified correlation with a criterion, however, depends on numerous other factors (e.g., manpower supply, validity of criteria, selection ratio). An analysis of these factors is beyond the scope of this research. However, it can be said that the correlations obtained between aptitude composite and SQT scores are of the magnitude of the highest obtained correlations for civilian employment tests (Ghiselli, 1966).

*Uncorrected correlation coefficients can be found in Table 25 of Appendix E.

Table 6

Correlations* of Aptitude Composite Scores
with SQT Performance for Eight Army MOSs

MOS	Composite	n	Percent of Tasks Go: Job-Site Component	Hands-on Component	Skill Component	Total SQT
11B	CO	24665	.04	.23	.55	.51
11C	CO	5806	.03	.25	.57	.50
19E	CO	4142	.07	.17	.57	.53
05C	SC	1737	.04	.14	.62	.56
31M	EL	2291	.01	.17	.56	.49
67N	MM	1394	.20	.31	.61	.62
73C	CL	634	.12	--**	.57	.54
75B	CL	477	.05	.34	.49	.52

* Corrected for restriction of range.
** Too few observations.

CORRELATIONS BETWEEN AFQT/APTITUDE COMPOSITE SCORES AND SQT PERFORMANCE FOR DIFFERENT RACIAL/ETHNIC GROUPS

The relationship between AFQT/aptitude composite scores and SQT performance for the three racial/ethnic groups is presented in Tables 7 and 8. These correlations are corrected for restrictions of range.

Findings

- Correlations between AFQT scores and SQT Skill Component scores average slightly higher for whites (.47) than for blacks (.39) or Hispanics (.40). This pattern is also evident in the correlations between AFQT scores and SQT Hands-On Component scores.

- An examination of the relationship between aptitude composite scores and SQT test scores reveals a similar pattern. For example, correlations between aptitude composite scores and Skill Component test scores average .52 for whites, .44 for blacks, and .49 for Hispanics.

Table 7

Correlations* Between AFQT Scores and Performance on Skill and Hands-On Components of the SQT for Different Racial/Ethnic Groups

		Percent of Tasks Go:					
		Skill Component			Hands-On Component		
	MOS	White	Black	Hispanic	White	Black	Hispanic
11B	Infantryman	.47	.40	.49	.16	.15	.12
11C	Indirect Fire Infantryman	.52	.34	.47	.18	.09	.16
19E	Armor Crewman	.47	.37	.35	.14	.13	.06
05C	Radio Teletypewriter Operator	.48	.39	.30	.12	.09	-.08
31M	Multichannel Communications Operator	.49	.30	.50	.12	.01	.31
67N	Utility Helicopter Repairer	.41	.55	.33	.16	.20	.03
73C	Finance Specialist	.41	.35	.46	-.07	-.23	.00**
75B	Personnel Administration Specialist	.50	.43	.28	.24	.17	.52

* Corrected for restriction of range.
** Too few observations.

Table 8

Correlations* Between Aptitude Composite Scores and Performance on Skill and Hands-On Components of the SQT for Different Racial/Ethnic Groups

			Percent of Tasks Go:					
			Skill Component			Hands-On Component		
	MOS	Aptitude Composite	White	Black	Hispanic	White	Black	Hispanic
11B	Infantryman	CO	.49	.46	.54	.22	.19	.20
11C	Indirect Fire Infantryman	CO	.54	.37	.49	.23	.25	.22
19E	Armor Crewman	CO	.53	.40	.51	.20	.18	-.01
05C	Radio Teletypewriter Operator	SC	.60	.50	.41	.15	.10	-.11
31M	Multichannel Communications Operator	EL	.53	.42	.60	.16	.16	.07
67N	Utility Helicopter Repairer	MM	.54	.58	.29	.27	.25	.28
73C	Finance Specialist	CL	.42	.48	.72	.20	-.12	.00**
75B	Personnel Administration Specialist	CL	.55	.30	.35	.36	.20	.48

* Corrected for restrictions of range.
** Too few observations.

SECTION III

RELATIONSHIP OF AFQT/ASVAB TO TRAINING PERFORMANCE MEASURES

This section of the report relates measures of training performance to scores on the AFQT and to scores on the aptitude composites.

CURRENT TRAINING PERFORMANCE CRITERIA

Current training performance criteria for the Marine Corps and Army occupational specialties included in this research are detailed in Table 9. In each entry-level training course, the current criterion, whether an overall percent correct or a minimum level of proficiency on a series of performance and/or written tests, can be expressed as an overall final course grade. Individual test grades are normally adjusted at each school when a trainee retakes a test and obtains an improved score. However, since this procedure can obscure real differences between individuals, the measures used here represent only the first attempts of trainees on each test.

Sample Characteristics

In order to determine whether samples selected for use in this research were representative of all accessions into each specialty, AFQT scores for obtained samples were compared to those for recent accessions. Data on accessions were supplied by the Defense Manpower Data Center and were only available for the Army MOSs (with the exception of Army MOS 19E, Armor Crewman). These data are displayed in Table 10. For the purpose of these comparisons, the data on the samples include both trainees who completed training and trainees who attrited from MOS training. The results indicate that school samples are representative of FY 1981 accessions. These comparisons were made with regard to the distribution of soldiers across AFQT categories for each MOS. Chi-square tests indicated that school samples and FY 1981 accessions did not differ significantly in terms of the proportion of individuals scoring at each AFQT level (alpha = .01).

The population of SQT takers, on the other hand, consisting of individuals who enlisted earlier than school samples or FY 1981 accessions, has a markedly different AFQT distribution (see Section II, Table 3). Whereas the proportion of Category IV personnel among SQT takers is over 50 percent, Category IVs account for less than 30 percent of school samples and FY 1981 accessions. Chi-square tests showed the proportion of SQT takers at each AFQT level was significantly different from that in the school samples and FY 1981 accessions.

Thus, school samples, but not SQT takers, are representative of recent accessions for the MOSs included in this study. The fact that each sample is representative of the total accessions in the MOSs included in this study is important because it allows us to generalize our findings from the sample of trainees in the study to the much larger group of recent accessions into the MOSs studied.

The training sample, which was drawn from four Marine Corps training courses (from two occupational specialties) and eight Army MOS training courses, was composed of 2,385 trainees. In this sample, 76 percent of the trainees were high school graduates and 24 percent were non-graduates.

Table 9

Course Organization and Current Standards for Course Completion in Two Marine Corps MOSs and Eight Army MOSs

Specialty	Course Length (weeks)	Course Organization	Current Standard
Marine Corps			
0311 Infantryman	3.6	lockstep	60% average on two written exams, physical fitness test, and commander's evaluation.
Army			
11B Infantryman	12.0	lockstep	80% of tasks must be passed on end-of-course performance test (POIQT)* (Additional requirements: qualify with M16 rifle and one other weapon, and meet PFT** requirement).
11C Indirect Fire Infantryman	12.0	lockstep	80% of tasks must be passed on end of course performance tests (POIQT) 70% or all non-zero marks on all parts of mortar of mortar qualification tests, meet PFT requirements.
19E Armor Crewman	13.0	lockstep	60% average on three gate performance tests.
05C Radio Teletypewriter Operator	10.8 (estimated)	self-paced	Must pass each of eleven module performance tests in order (including 25 wpm typing, International Morse Code). Standards

* Performance Oriented Infantry Qualification Test
** Physical Fitness Test

Table 9 continued

Specialty	Course Length (weeks)	Course Organization	Current Standard
			are fixed but vary across modules. Must also pass comprehensive end-of-course test (chance to retest and change No Go's to Go's).
31M Multichannel Communications Operator	7.0	lockstep	70% on each of six performance tests.
2841 Ground Radio Repair			
Basic Electronics (BEC)	12.2	lockstep	70% average on three performance and eleven written tests.
Radio Fundamentals (RFC)	4.2	lockstep	70% weighted average based on four practical tests (40% weight), fourteen quizzes (20% weight), and three written exams (40% weight).
Ground Radio Repair (GRRC)	12.2	lockstep	70% weighted average based on ten performance tests (45.5%), ten written tests (28.5%), and 32 quizzes (26%).
67N Utility Helicopter Repairer	10.06 (estimated)	self-paced	70% on each of eight performance tests and nine written tests.
73C Finance Specialist	7.0	lockstep	70% on each of six written performance tests.
75B Personnel Administration Specialist	7.4 (estimated)	self-paced	Must pass fourteen of fifteen written performance tests (including 20 wpm typing). Standards vary from 80-100% on these tests.

Table 10

Distribution of AFQT Scores for FY 1981 Accessions and for Training Samples from Eight Army MOSs

MOS	Group	n	(Percent) I	II	IIIA	IIIB	IVA	IVB
11B	Accessions (81)	4,083	3	24	17	31	13	12
	School Sample	195	3	29	17	31	10	11
11C	Accessions (81)	1,044	2	17	14	33	21	13
	School Sample	178	2	19	18	32	2	9
19E	Accessions (81)	---	-	--	--	--	--	--
	School Sample	220	2	21	17	28	16	17
05C	Accessions (81)	1,201	2	33	26	31	6	1
	School Sample	200	2	29	26	32	9	5
31M	Accessions (81)	694	1	25	22	33	10	10
	School Sample	222	1	22	20	25	16	16
67N	Accessions (81)	486	5	41	23	24	4	3
	School Sample	168	7	42	22	21	4	4
73C	Accessions (81)	279	5	25	15	27	20	8
	School Sample	188	4	25	14	27	21	9
75B	Accessions (81)	425	2	18	16	33	22	9
	School Sample	162	2	17	13	33	23	12

RELATIONSHIP BETWEEN FINAL COURSE GRADES AND AFQT/ASVAB SCORES

Figures 51 to 62 in Appendix F display, for each of the 12 entry-level training courses in this study, final course grades as a function of AFQT scores separately for high school and non-high school graduates. Figures 63 to 74 in Appendix F similarly display final course grades as a function of aptitude composite scores separately for high school and non-high school graduates. Final course grades are expressed in these figures on a scale of 100. When insufficient data were available to reliably determine a point (i.e., fewer than six trainees), that point was omitted from the figure. (This occurred frequently in specialties with small numbers of non-high school graduates.) The correlations between final course grades and AFQT/ASVAB scores are presented later in this section in Tables 14 and 15.

The samples included in this analysis are unrepresentative of enlistees in general in at least two ways. First, individuals who fail to complete a course do not receive a final course grade and therefore are not included in this data set. Second, minimum scores are required on various aptitude composites in order to qualify for each specialty. Since both attrition rates and classification requirements varied across specialties, the range of abilities found among the course completers in the specialties studied was quite variable.

Findings

An examination of Figures 50 through 74 reveals that the relationship between AFQT scores and final course grades parallels closely the relationship between aptitude composite scores and final course grades. Therefore, these data will be discussed together.

- Final course grades were very high across all specialties, with mean grades ranging from 81 percent to 94 percent. (Acceptable grades are detailed for each specialty in Table 9.) There is a positive relationship between AFQT/ASVAB scores and final course grades for all specialties. Final course grades for trainees with the highest scores on AFQT/ASVAB for their specialty were generally from five to 10 percentage points higher than those for trainees with the lowest AFQT/ASVAB scores for their specialty.

- The relationship between final course grades and AFQT/ASVAB appeared to be weaker for combat arms specialties (e.g., Army and Marine Corps Infantryman) than for technical (e.g., Utility Helicopter Repairer) or administrative (e.g., Finance Specialist) specialties. The simplest explanation for these findings is that combat arms training is less academically demanding than training in technical or administrative specialties. Therefore, ASVAB may have less utility in predicting success for combat arms than for technical or administrative specialties.

- Final course grades were slightly higher for high school graduates than for non-graduates with similar AFQT/ASVAB scores. The advantage of high school graduates over non-graduates was slightly greater in technical (e.g., for 67N, Utility Helicopter Repairer, four percentage points) or administrative (e.g., 75B, Personnel Administration Specialist,

three percentage points) specialties than in combat arms (e.g., 11B, Infantryman, one percentage point; 0311, Infantryman, one percentage point). The fact that trainees who failed to complete their high school education had more difficulty attaining high grades in entry-level military training courses, particularly more technical courses, should not be surprising. High school graduates have advantages in terms of specific skills and knowledge, and also may be more stable or mature than non-high school graduates. In addition, statistical regression effects may have contributed to non-graduates' lower course grades. The AFQT/ASVAB scores of non-high school graduate enlistees are not representative of those of non-graduates in general because selection standards disqualify non-graduates with low scores. Non-graduate enlistees' "true" AFQT/ASVAB scores are lower (closer to the non-graduate group mean) than their observed scores, and hence their scores on any variable correlated with AFQT/ASVAB would tend to regress toward the non-graduate group mean.

- The pattern of final course grades as a function of AFQT/ASVAB scores was very similar for high school and non-high school graduates. In some specialties, however, there was an insufficient number of non-high school graduates to make a definite determination.

RELATIONSHIP BETWEEN FINAL COURSE GRADES AND AFQT/ASVAB SCORES FOR DIFFERENT RACIAL/ETHNIC GROUPS

The samples of trainees from eight Army and four Marine Corps training courses for which final course grades were obtained were comprised of 72 percent whites, 21 percent blacks, 4 percent Hispanics, and 3 percent spread among other racial/ethnic groups (e.g., American Indian, Filipino).

Figures 75 to 82 contain data only on those occupational specialties in which at least two racial/ethnic groups each comprised 20 percent or more of the sample. The four specialties which met this requirement were Radio Teletypewriter Operator (05C), Multichannel Communications Operator (31M), Finance Specialist (73C), and Personnel Administration Specialist (75B). In each case, only the white and black racial/ethnic groups comprised a large enough proportion of the sample to be included. Figures 75 to 78 display final course grades as a function of AFQT scores and racial/ethnic group. Figures 79 to 82 display final course grades as a function of aptitude composite scores and racial/ethnic group. Tables 16 and 17 later in this section contain correlation coefficients between AFQT/ASVAB scores and measures of training performance for the differenct racial/ethnic groups. (Uncorrected correlation coefficients are in Tables 30 and 31 of Appendix I.)

Findings

The relationship of AFQT scores to final course grades for each racial/ethnic group displayed in the figures located in Appendix F parallels closely the relationship between aptitude composite scores and final course grades for the same racial/ethnic groups.

- Final course grades increased as a function of increasing AFQT/ASVAB scores for both the white and black groups. These increases in final course grades across the range of AFQT/ASVAB scores varied from five to 15 percentage points for both racial/ethnic groups.

- Members of the two racial/ethnic groups who had the same AFQT/ASVAB scores generally averaged the same final course grades. This finding is in contrast to the finding in Section II that AFQT/ASVAB scores predicted higher SQT scores than were actually achieved for blacks.

ALTERNATIVE TRAINING CRITERIA

One purpose of this research was to examine the relationship between current criteria for course completion and AFQT/ASVAB scores. A second goal was to identify other existing criteria and to develop new criteria which might have potential value. Therefore, the first step in this part of the report will be to summarize alternative existing criteria. Second, data on the criteria developed as part of this research will be presented.

ALTERNATIVE EXISTING TRAINING DATA

Attrition

Attrition is an important performance measure in a training course. Attrition is variously labeled as academic, administrative, motivational, medical, etc. In fact, these distinctions are difficult to make in many cases. Therefore, for the purposes of this research, all types of attrition were combined.

Attrition rates varied significantly across occupational specialties. Training courses, whether difficult or easy, may produce high or low attrition depending on the required standards of performance required for course completion.

The trainees who enter a skill training course have been pre-screened in two ways which reduce the likelihood that they will fail to complete training:

- In most specialties, trainees have completed basic training, a process which eliminates many who do not have the ability and motivation to be trained in a skill.

- The requirement for minimum scores on specific aptitude composites insures that most students can cope with the subject matter of the course they attend.

In addition, some training courses are self-paced, allowing slower trainees to get more attention and more time to complete training (e.g., 05C, Radio Teletypewriter Operator). Even in some courses which are not self-paced, provisions have been made for students with unsatisfactory performance to be "recycled", that is, to retake a module or part of a course (e.g., 2841, Marine Corps Ground Radio Repair-Basic Electronics Course).

The overall attrition rate in the training courses studied was 12 percent. This finding is consistent with attrition rates between 8 and 11 percent reported from 35 Army skill training courses during the period from 1976 to 1979 (Greenberg, 1980).

The problem of attrition does not end for the Army or Marine Corps after completion of skill training. Nearly 40 percent of all enlistees in the Army and Marine Corps fail to complete their initial three-year enlistment (Sinaiko & Schefien, 1980). The causes for attrition have been hypothesized to include:

- o Characteristics of the individual who is discharged. These characteristics include adaptability to a military environment and possession of needed skills and knowledge.

- o Characteristics related to the organization and its policies and practices. Attrition varies significantly across units and specialties, suggesting that pay, opportunities for promotion, leadership, and so on have an impact on attrition (Goodstadt & Yedlin, 1979).

In order to examine the relationship between attrition from entry-level training courses and those characteristics listed above, attrition patterns were examined for all courses in the study with attrition rates of 10 percent or more (i.e., 05C, Radio Teletypewriter Operator; 2841, Marine Corps Ground Radio Repair-Basic Electronics Course; 19E, Armor Crewman; and 31M Multichannel Communications Operator).

High school graduation might be negatively related to attrition because it may reflect, in addition to cognitive skills, the ability to adapt to a disciplined environment. Such an ability would seem important for success in the military. Differences in attrition rates from entry-level training courses for high school graduates and non high-school graduates are shown below in Table 11:

Table 11

Attrition Rates as a Function of Level of Education

Specialty Entry-Level Training Course		Proportion of Trainees Failing to Complete Training	
		Non-High School Graduates	High School Graduates
19E	Armor Crewman	.22	.15
05C	Radio-Teletypewriter Operator	.39	.21
31M	Multichannel Communications Operator	.24	.16
2841	Marine Corps Ground Radio Repair-Basic Electronics Course	.58	.32

Thus, high school graduates are less likely to drop out of entry-level training. This finding is especially impressive in light of the fact that high school graduate enlistees did not have higher AFQT scores than non-high school graduate enlistees in these specialties, as shown below in Table 12.

Table 12

Mean AFQT Scores as a Function of Level of Education

Specialty Entry-Level Training Course		Mean AFQT Score	
		Non-High School Graduates	High School Graduates
19E	Armor Crewman	44.7	44.7
05C	Radio Teletypewriter Operator	49.8	53.0
31M	Multichannel Communications Operator	52.6	42.0
2841	Marine Corps Ground Radio Repair - Basic Electronics Course	70.0	72.2

Since success in entry-level training requires significant cognitive skills, it is also reasonable to expect that attrition from such training courses is related to measures of such skills and abilities. Thus, attrition can be expected to vary with AFQT/ASVAB scores. The data for the qualifying training courses are displayed in Figures 83 and 84 in Appendix F. The percent attrited is presented as a function of AFQT categories in Figure 83 and as a function of aptitude composite scores in Figure 84.

Findings

The relationship between AFQT scores and attrition rates parallels closely the relationship between aptitude composite scores and attrition rates. Therefore, these data will be discussed together.

- There is an inverse relationship between AFQT/ASVAB scores and attrition rates for two specialties (2841 Marine Corps Ground Radio Repair-Basic Electronics Course; 05C Radio Teletypewriter Operator) and no relationship for two other specialties (19E, Armor Crewman; 31M, Multichannel Communications Operator).

- In the Marine Corps Ground Radio Repair-Basic Electronics Course, attrition is at a relatively low rate of 20 percent for the trainees in the two highest AFQT categories (or highest aptitude composite category) but rises to over 60 percent for all lower-scoring trainees. The Basic

Electronics Course (BEC) is the first of three courses which make up initial entry training for the Ground Radio Repair specialty. The BEC is a theory-based course. As such, it may represent something of an artificial hurdle for soldiers entering the Ground Radio Repair specialty.

Alternatively, the criteria for accession into the Ground Radio Repair Course may be too low, allowing unqualified personnel into the specialty.

- In the Army Radio Teletypewriter Operator (05C) course attrition increased steadily from about 20 to 40 percent as a function of decreasing AFQT/ASVAB scores.

Other studies have obtained data linking attrition to performance on cognitive tests such as the AFQT/ASVAB (Lockman, 1977; Matthews, 1977; Sinaiko & Scheflen, 1980; Dann, 1978). However, requirements for accession into a specialty are intended to insure that all individuals are capable of completing training. Furthermore, many training courses are flexible, allowing for differences in course completion time. Consequently, attrition should be primarily an adaptability criterion, relatively free from cognitive effects. The fact that high rates of attrition do occur and show a relationship with AFQT/ASVAB scores even for samples in which every soldier is presumed to possess the necessary intellectual ability suggests either

- excessive standards of performance in training courses; or
- insufficiently high standards for selection and classification of soldiers into a specialty.

Either situation would result in the kind of relationship observed between attrition and AFQT/ASVAB scores for two specialties in this study, 2841 (Marine Corps Ground Radio Repair-Basic Electronics Course) and 05C (Radio Teletypewriter Operator).

Time-to-Complete

The time required to complete a training course, especially one which is self-paced, is generally considered indicative of the ability and motivation of a trainee. Some courses which are not self-paced allow slower learners extra training time by letting them repeat modules or components of a course. In either case, time-to-complete (TC) is potentially a useful measure because those who take longer to complete their training consume additional training resources and are not available for service in units until training is completed. Of course, time-to-complete measures also have some problems as criteria. For example, in some training courses soldiers who finish early must wait until other soldiers in their starting group have completed the course before receiving their next assignment. Such situations reduce the motivation of soldiers to complete a course as rapidly as possible. In such a system, there are no rewards for soldiers who complete training quickly.

Data for those training courses in this study which are self-paced (i.e., 05C, Radio Teletypewriter Operator; 67N, Utility Helicopter Repairer; 75B, Personnel Administration Specialist) or allow trainees to repeat unsatisfactory work (i.e.,

2841, Marine Corps Ground Radio Repair-Basic Electronics Course), resulting in a distribution of times to complete, are presented in Figures 85 and 86 in Appendix F. Time-to-complete (TC) is displayed as a function of AFQT score in Figure 85 and as a function of aptitude composite score in Figure 86. Time-to-complete is expressed as the number of training days (based on a five-day week) for soldiers with particular AFQT/ASVAB scores to complete their training course. The correlations between time-to-complete and AFQT/ASVAB scores are presented in Tables 14 and 15 and are discussed later in this section of the report.

Findings

It is clear that both AFQT and aptitude composite scores relate to time-to-complete in a parallel fashion. Consequently, these data will be discussed together.

- There is an inverse relationship between AFQT/ASVAB scores and time to complete training for all four specialties. The difference in training time required by trainees in the lowest and highest AFQT/ASVAB score categories was comparable for all specialties, averaging approximately 12 training days. Thus, for every 10-point difference in aptitude composite scores, one can expect trainees with lower scores to require about 3 more days of training.

Alternative Performance Tests

In some occupational specialties, performance tests which are not part of the final course grade are administered on a pass/fail basis. Typically, the scoring on these tests does not distinguish among varying levels of proficiency. However, the Mortar Qualification Test (MQ) administered to Indirect Fire Infantrymen (11C), does make distinctions among qualified individuals and therefore has potential as a criterion. Figures 87 and 88 contain data on MQ test scores as a function of AFQT/ASVAB performance and level of education (see Tables 14 and 15 for correlations).

Findings

The data relating both AFQT (Figure 87) and aptitude composite scores (Figure 88) to MQ test scores are similar and therefore will be discussed together.

- MQ test performance is directly related to AFQT/ASVAB scores. MQ test scores increase about 10 percentage points over the full range of AFQT/ASVAB scores.

- There were no differences between high school and non-high school graduates in either overall level of performance on the MQ or in the relationship between MQ test performance and AFQT/ASVAB scores.

ALTERNATIVE EXPERIMENTAL TRAINING CRITERIA

Two types of experimental training measures were used in this research. The first measure consisted of peer nominations on a number of dimensions reflecting attitudes or abilities not measured by currently administered training performance

measures. These measures were obtained for four occupational specialties as shown in Table 13.

Table 13

Peer Nomination Samples

Service	Specialty		Sample Size
Marine Corps	0311	Infantryman	132
Army	11B	Infantryman	115
Army	31M	Multichannel Communications Operator	131
Army	73C	Finance Specialist	58

In order to complete a peer nomination form, each trainee was required to designate six individuals in his class or group, other than himself, who best fit each of a series of 12 descriptions (e.g., highly motivated, tries hard to succeed in training). These descriptions covered six dimensions (motivation, ability to communicate, leadership ability, proficiency with equipment, cooperativeness, and soldiering) with both a negative and a positive description for each (e.g., highly motivated, lacks motivation). An individual's score on any dimension was obtained by separately summing positive and negative attributions, subtracting negatives from positives, and normalizing the resulting score with respect to the size of the rating group. This procedure yields dimension scores for each individual which may vary from -1 to +1.

A second experimental measure of training performance was an instructor/supervisor rating instrument field tested in the Army MOS 11B, Infantryman (administered on 115 soldiers). Each instructor/supervisor was asked to rate every trainee in a class/group on each of 10 dimensions reflecting attitudes or abilities (e.g., dependability, motivation, ability to communicate). Responses were made on a seven-point scale.

A brief literature review of issues related to the subjective performance instruments developed for this research is provided in Appendix G. This is followed, in Appendix H, by detailed information on these experimental training criteria including

- o specific criteria and procedures for administering the instruments; and
- o copies of the experimental instruments.

Peer Nominations

Peer nomination data on each of the six dimensions rated were examined as a function of AFQT/ASVAB scores. All dimensions showed a similar relationship to AFQT/ASVAB scores. This result can be explained by the fact that the peer

attribution scores on the six dimensions were highly related: intercorrelations among the six dimension scores within various specialties averaged about .80. Thus, trainees receiving high attribution scores on one dimension tended to receive them also on other dimensions. This conclusion is confirmed by factor analyses which were run on the peer nomination data for each specialty. For each specialty, the factor analyses showed that there is actually one factor underlying the six peer dimensions and that all six of the peer dimensions contributed equally to that factor.

For illustrative purposes, Figures 89-96 in Appendix F show the relation of the peer nomination dimension "ability to communicate" to AFQT and aptitude composite scores. Although this dimension is not fully representative of the particular underlying factor common to all six dimensions, it did have a slightly stronger relationship to AFQT/ASVAB than the other peer nomination dimensions.

For each of the four specialties (Marine Corp Infantryman, 0311; Army Infantryman, 11B; Multichannel Communications Operator, 31M; Finance Specialist, 73C) with peer nomination data, peer scores on the communication dimension are presented as a function of AFQT scores and level of education in Figures 89 to 92 in Appendix F and as a function of aptitude composite score and level of education in Figures 93 to 96 (see Tables 14 and 15 for correlations).

Findings

The relationship between AFQT scores and peer scores paralleled closely the relationship between aptitude composite scores and peer scores. As a result, these data will be discussed together.

- The mean peer rating score for each specialty was basically fixed, because for each positive rating given (coded as +1) a negative rating was also required (coded as -1). As a result, mean ratings approximated zero for all dimensions in all specialties (ranged between -.03 and .06). The standard deviations of peer ratings across rated dimensions within each specialty were quite uniform, varying less than .05 from highest to lowest. While uniform within each specialty, standard deviations of ratings across individuals did vary from one specialty to another (i.e., MC 0311, SD=.32; 11B, SD=.46; 31M, SD=.20; 73C, SD=.39). The meaning of these standard deviations can be illustrated with an example. If, for instance, members of a peer group, having been assembled for the first time, were asked to rate one other, ratings would in effect be random and no individuals would be expected to be distinguished either as outstanding or inept. As a result, the standard deviation of mean ratings across individuals in this group would be relatively small. If, on the other hand, ratings were obtained on a peer group who had been through several months of intensive training together, a consensus is likely regarding outstanding and inept performers, resulting in a much higher standard deviation of mean ratings across individuals. Therefore, these standard deviations for peer ratings can be interpreted as an agreement index; that is, a high standard deviation is indicative that a consensus exists regarding an individual's deserved rating (positive or negative).

- There is a direct relationship between AFQT/ASVAB scores and peer nomination scores for all four specialties, but the strength of the

relationship varies (correlations can be found in Tables 14 and 15). The increase in peer nomination scores over the range of AFQT/ASVAB scores varies from .10 for Multichannel Communications Operator (31M) to .70 for both Infantryman (11B) and for Finance Specialist (73C).

- The relationship between AFQT/ASVAB scores and peer nomination scores was similar for high school and non-high school graduates. The peer scores of high school graduates, however, averaged .10 to .20 points higher than those of non-high school graduates across the range of AFQT/ASVAB scores.

Instructor Ratings

Instructor ratings were administered in only one occupational specialty, Infantryman (11B). The ratings were made on 10 dimensions (e.g., motivation, dependability). Since the data on all dimensions were similar and since intercorrelations of dimensions were uniformly high (i.e., averaged over .70), only one dimension, the ability to work with equipment, is presented here. Instructor ratings are displayed in Figure 97 as a function of AFQT scores and level of education and in Figure 98 as a function of aptitude composite scores and education. (These figures are both in Appendix F.)

Findings

The relationship of instructor ratings to AFQT and aptitude composite scores was similar, so these data will be discussed together.

- The mean rating on each dimension ranged from 4.71 to 5.43. The standard deviations were quite uniform, ranging from 1.36 to 1.77.

- There is a positive relationship between AFQT/ASVAB scores and instructor ratings of the ability of soldiers to work with equipment. Instructor ratings for soldiers in the highest AFQT/ASVAB categories were one point higher on a seven-point scale than ratings for soldiers in the two lowest AFQT/ASVAB categories, a statistically significant difference, $F(1,109) = 18.11$, $p \leq .01$.

- The relationship between AFQT/ASVAB scores and instructor ratings was comparable for high school and non-high school graduates. The instructor ratings of high school graduates, however, were about one point higher than those for non-high school graduates with the same AFQT/ASVAB scores.

CORRELATIONS BETWEEN AFQT AND TRAINING PERFORMANCE MEASURES

The correlations between AFQT and various criteria of training performance are displayed in Table 14. A correlation between AFQT scores and training criteria in a specialty is an estimate of the validity of AFQT as a predictor of success in training in that specialty, but also depends upon the reliability of the training measure. The resulting correlations must be corrected for range restriction in the

training samples caused by:

- o use of AFQT as a selector into the military
- o use of aptitude composites as selectors into occupational specialties
- o attrition which takes place before training criteria are obtained
- o a lack of high scoring (on AFQT/ASVAB) trainees in many training courses.

The correlations presented in Table 14 are corrected for restricted range resulting from those factors listed above. (Uncorrected correlations can be found in Table 28 in Appendix I.)

Findings

There are several trends evident in the data.

- o Correlations between AFQT scores and final course grades range from .12 to .68. These correlations seem to vary systematically. That is, for combat arms specialties, the correlations are low to moderate (i.e., between .12 and .38) while for technical (i.e., communications operation and maintenance, mechanical equipment repair) and administrative specialties the correlations are moderate to high (i.e., between .33 and .68). These findings are consistent with those of Valentine (1977), who found, for 43 Air Force training courses, uncorrected correlations for non-technical courses ranging up to .40 (as compared to .38 in this study) and for technical courses ranging up to .49 (as compared to .68 in this study).
- o The magnitude of correlations between attrition and AFQT scores with the exception of Marine Corps 2841 BEC (r = -.59), are quite low (-.06 to -.21).
- o Correlations between AFQT scores and time-to-complete (TC) measures are moderate, ranging from -.23 to -.49. These correlations are somewhat less than correlations between AFQT scores and final course grades. These lower correlations may be due to some of the factors mentioned earlier which contaminate time-to-complete indices.

Table 14

Correlations* Between AFQT Scores and Measures of Training Performance

	Specialty	Final Course Grade	Attrition**	Time To Complete	Peer Nomination	Instructor Rating	Alternate Performance Measures
0311	Infantryman	.24			.40		
11B	Infantryman	.23			.45	.23	
11C	Indirect Fire Infantryman	.12					.21
19E	Armor Crewman	.38	-.10				
05C	Radio Teletypewriter Operator	.33	-.21	-.27			
31M	Multichannel Communications Operator	.33	-.06		.12		
2841	Basic Electronics	.48	-.59	-.23			
2841	Radio Fundamentals	.39					
2841	Ground Radio Repair	.36					
67N	Utility Helicopter Repairer	.68		-.49			
73C	Finance Specialist	.39			.58		
75B	Personnel Administration Specialist	.36		-.32			

* Correlations are corrected for restriction in range

** Because attrition is a dichotomous variable, the resulting biserial correlation coefficients cannot be translated to Pearson product-moment coefficients. Thus, correlations shown underestimate the relationship between attrition and AFQT scores.

- Correlations between AFQT scores and peer nomination scores are moderate to high for three specialties, ranging from .40 to .58. In the fourth specialty, Multichannel Communications Operator (31M), the correlation is low (.12). This low correlation may be explained by the fact that the trainees who participated in the peer nomination process from 31M had been together for a shorter period of time and were in larger peer groups (average group size about 28) than the other specialties (average group size about 20) which participated. This could have reduced the reliability of the nominations obtained and, consequently, the correlation (see Appendix G). The correlations between AFQT scores and peer nomination scores for the other three specialties exceed the corresponding correlations between AFQT scores and final course grades by between .16 and .22. Thus, for three of four occupational specialties, peer nomination scores appear to be more strongly related to AFQT than average final course grades. Despite the encouraging results found here for peer nomination procedures, it should be remembered that a number of difficulties often attend attempts to use such procedures in operational settings. These limitations are discussed in some detail in Appendix G.

- The correlation between AFQT scores and instructor ratings for Infantryman (11B) is .23, equal to the correlation between AFQT scores and final course grades. Once again, rating systems have limitations in their application in operational settings. These are discussed in Appendix G.

CORRELATIONS BETWEEN APTITUDE COMPOSITES AND TRAINING PERFORMANCE MEASURES

The correlations between aptitude composite scores and training performance measures are displayed in Table 15. Aptitude composite scores of the ASVAB are used to predict specialties in which an enlistee has a high probability of success. Since training is the first 'gate' in a specialty through which recruits must pass, aptitude composite scores should be correlated with measures of performance in training. However, because the soldiers in any specialty have been selected on the basis of aptitude composite scores, there is a restriction in range of aptitude composite scores. Therefore, all reported correlations are corrected for this restriction (uncorrected correlations are in Table 29 in Appendix I).

Findings

These data resemble the correlations between AFQT scores and training performance measures.

- Correlations between aptitude composite scores and final course grades range from .21 to .79. As before, in combat arms specialties, correlations are low to moderate (i.e., between .21 and .51) while for technical or administrative specialties, correlations are high (i.e., between .44 and .79). These findings are consistent with other research (Hiatt & Sims, 1980; Vitola, Mullins, & Croll, 1973) in which the uncorrected correlations between aptitude composites and final course grades are around .30

Table 15

Correlations* Between Aptitude Composite Scores and Measures of Training Performance

	Occupational Specialty	Composite	Final Course Grade	Attrition**	Time to Complete	Peer Nomination	Instructor Rating	Alternate Performance Measures
0311	Infantryman	CO	.36			.41		
11B	Infantryman	CO	.26			.40	.26	
11C	Indirect Fire Infantryman	CO	.21					.34
19E	Armor Crewman	CO	.51	-.20				
05C	Radio Teletypewriter Operator	SC	.44	-.15	-.38			
31M	Multichannel Communications Operator	EL	.68	-.35		.35		
2841	Basic Electronics	EL	.76	-.68	-.57			
2841	Radio Fundamentals	EL	.74					
2841	Ground Radio Repair	EL	.50					
67N	Utility Helicopter Repairer	MM	.79		-.63			
73C	Finance Specialist	CL	.63			.80		
75B	Personnel Administration Specialist	CL	.44		-.48			

* Correlations are corrected for restriction of range.

** Because attrition is a dichotomous variable, the resulting biserial correlation coefficients cannot be translated to Pearson product-moment coefficients. Thus, correlations shown underestimate the relationship between attrition and aptitude composite scores.

(averaged around .25 in this study) for combat arms, while corresponding correlation coefficients for mechanical maintenance and electronics specialties are around .55 (averaged .50 in this study).

- Correlations between attrition and aptitude composite scores, with the exception of Marine Corps 2841 BEC (r = -.68), are fairly low (-.15 to -.35).

- Correlations between aptitude composite scores and time-to-complete measures are moderate to high, ranging from -.38 to -.63.

- Correlations between aptitude composite scores and peer nomination scores are moderate to high, ranging from .35 to .80. Once again, the correlation for Multichannel Communications Operator may have been reduced by unreliability in the ratings, accounting for the higher correlation between aptitude composite and final course grade (.68 vs. .35). For the other three specialties, however, the correlations between aptitude composite scores and peer nomination scores exceed corresponding correlations between aptitude composites and final course grades.

- The correlation between aptitude composite scores and instructor ratings for Infantryman (11B) is .26, equal to the correlation between aptitude composite scores and final course grades.

CORRELATIONS BETWEEN AFQT SCORES AND TRAINING PERFORMANCE MEASURES FOR DIFFERENT RACIAL/ETHNIC GROUPS

The correlations between AFQT and various measures of training performance are displayed in Table 16. All reported correlations are corrected for restriction of range. (Uncorrected correlations are contained in Table 30 of Appendix I.) For each of the four occupational specialties displayed, there were too few (<30) Hispanics to compute correlations. Therefore, the correlation coefficients presented in Table 16 reflect the relationship between AFQT scores and training performance for whites and blacks only.

Findings

- Correlations between AFQT scores and final course grades are moderate, ranging from .26 to .51, and are similar for whites and blacks (differing by .06 or less), with the exception of Multichannel Communications Operator, in which the correlation was much higher for whites than for blacks (.51 vs. .05).

- For both MOSs in which there is significant attrition (05C and 31M), correlations between AFQT scores and attrition are higher for whites than for blacks. (For MOS 05, the correlation is -.38 for whites and -.04 for blacks. Similarly, for MOS 31M, correlations for whites and blacks are -.18 and .02, respectively.) This difference in correlations may be partially attributed to the fact that overall attrition rates in these MOSs are higher for whites than for blacks. For example, in MOS 05C, the

Table 16

Correlations* Between AFQT Scores and Measures of Training Performance Across Racial/Ethnic Groups for Four MOSs

		Final Course Grade			Attrition**		
	MOS	White	Black	Hispanic	White	Black	Hispanic
05C	Radio Teletypewriter Operator	.26	.32	***	-.38	-.04	***
31M	Multichannel Communications Operator	.51	.05	***	-.18	.02	***
73C	Finance Specialist	.39	.43	***			
75B	Personnel Administration Specialist	.38	.42	***			

		Time-to-Complete			Peer Nomination		
		White	Black	Hispanic	White	Black	Hispanic
05C	Radio Teletypewriter Operator	-.01	-.43	***			
31M	Multichannel Communications Operator				.23	.22	***
73C	Finance Specialist				***	***	***
75B	Personnel Administration Specialist	-.38	-.41	***	***	***	***

* Corrected for restrictions in range.

** Because attrition is a dichotomous variable, the resulting biserial correlation coefficients cannot be translated to Pearson product-moment coefficients. Thus, correlations should underestimate the relationship between attrition and AFQT scores.

*** Sample size too small (N<30).

mean attrition rate is 27 percent for whites and 17 percent for blacks. Likewise, in MOS 31M, the mean attrition rate for whites and blacks is 21 percent and 13 percent, respectively. Thus, while non-cognitive factors may account for some attrition in each MOS for blacks and whites, other attrition among whites is related to AFQT scores.

- The magnitude of the correlation between AFQT scores and time-to-complete for 05C (Radio Teletypewriter Operator) is higher for blacks than for whites (i.e., -.43 vs. -.01). There is no difference between the correlations for blacks and whites in MOS 75B.

- The correlations between AFQT and peer nomination scores are the same for blacks and whites for Multichannel Communications Operator 31M.

CORRELATIONS BETWEEN APTITUDE COMPOSITE SCORES AND TRAINING PERFORMANCE MEASURES FOR DIFFERENT RACIAL/ETHNIC GROUPS

The correlations between aptitude composite scores and training performance measures are displayed in Table 17. All reported correlations are corrected for restriction of range. As previously stated, there were too few Hispanics (<30) to compute correlations; therefore, the correlations coefficients for the four occupational specialties displayed in Table 17 are for whites and blacks only.

Findings

- Correlations between aptitude composite scores and final course grades are similar for whites and blacks, averaging .52 and .47 respectively.

- Correlations between aptitude composite scores and attrition are similar for whites (-.22) and blacks (-.25) for Radio Teletypewriter Operator. However, for Multichannel Communications Operator, the relationship is stronger for blacks than for whites (-.51 vs. -.28).

- Correlations between aptitude composite scores and time-to-complete are slightly stronger for blacks than for whites in both MOS 05C (-.42 vs. -.18) and MOS 75B (-.64 vs. -.52).

- The correlation between aptitude composite scores and peer nomination scores for Multichannel Communications Operator is slightly stronger for blacks (.65) than for whites (.30).

Table 17

Correlations* Between Aptitude Composite Scores and Measures of Training Performance Across Racial/Ethnic Groups for Four MOSs

	MOS	Aptitude Composite	Final Course Grade			Attrition**		
			White	Black	Hispanic	White	Black	Hispanic
05C	Radio Teletypewriter Operator	SC	.24	.49	***	-.22	-.25	***
31M	Multichannel Communications Operator	EL	.78	.43	***	-.28	-.51	***
73C	Finance Specialist	CL	.66	.56	***			
75B	Personnel Administration Specialist	CL	.38	.38	***			

	MOS	Aptitude Composite	Time-to-Complete			Peer Nomination		
			White	Black	Hispanic	White	Black	Hispanic
05C	Radio Teletypewriter Operator	SC	-.18	-.42	***			
31M	Multichannel Communications Operator	EL				.30	.65	***
73C	Finance Specialist	CL				***	***	***
75B	Personnel Administration Specialist	CL	-.52	-.64	***	***	***	***

* Corrected for restrictions in range.

** Because attrition is a dichotomous variable, the resulting biserial correlation coefficients cannot be translated to Pearson product-moment coefficients. Thus correlations shown underestimate the relationship between attrition and aptitude composite scores.

*** Sample size too small (N<30).

SECTION IV

ASSESSMENT OF JOB PERFORMANCE MEASURES

This chapter of the report addresses the adequacy of existing job performance measures as criteria for validation of AFQT/ASVAB measures. This chapter begins with a brief description of the ASVAB. In subsequent sections the Army job analysis program and entry-level training courses are discussed as they relate to job performance measurement. Finally, the SQT program which provides the principal measures in the Army of individual job proficiency, is reviewed.

ARMED SERVICES VOCATIONAL APTITUDE BATTERY

The ASVAB is a battery of tests which is used primarily to make selection and classification decisions for applicants to the Armed Services.

For the period covered by this study, 1976 to 1981, two versions of the ASVAB were in use. First, from January 1976 through September 1980, ASVAB test forms 6 and 7 were in effect. This ASVAB was composed of 13 subtests, listed in Appendix A. Its AFQT composite consisted of three subtests (Word Knowledge, Arithmetic Reasoning, and Space Perception). Other combinations of these subtests, called aptitude composites, were developed by each of the Services, in order to predict success in training for various occupational specialties. Those aptitude composites derived from ASVAB forms 6 and 7 which are relevant to the specialties in this study are also summarized in Appendix A.

In October 1980, new versions of the ASVAB were introduced, forms 8, 9 and 10, composed of 10 subtests. The AFQT was now composed of four subtests: Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Numerical Operations. A new set of aptitude composites were devised by the Services. The subtests and composites for ASVAB forms 8, 9, and 10 are summarized in Appendix A.

AFQT scores are reported as percentile scores and are used as the primary selection criterion for enlistment. In addition, enlistment eligibility is also determined by aptitude composite scores, which in the Army and Marine Corps are reported as standard scores, with a mean of 100 and a standard deviation of about 18. Finally, level of educational attainment (e.g., non-high school graduate, GED, or high school graduate) is an additional criterion in the selection process. The selection criteria for both the Army and Marine Corps are outlined in Appendix J.

In a personnel management system, a selection and classification test has several important functions. Most importantly, a selection test must be predictive of job performance. This can be achieved by conducting a careful analysis of job tasks, conditions, and standards for various levels of competency and by basing the development of a selection test on this analysis. These antecedent steps serve to add confidence that correlations between selection test scores and job performance measures represent the actual relationship between selection test scores and job performance.

JOB ANALYSIS

The principal measure of proficiency on specified tasks for soldiers in the Army is the SQT. The SQT is officially described (U.S. Army Training and Doctrine Command Regulation and Pamphlet 351-2) as a major training diagnostic for the individual training system. Thus, the primary purpose of SQTs is not to evaluate individual soldier performance but, rather, to diagnose individual training needs and to evaluate training program effectiveness.

The foundation for the SQT is the Army's individual training system. The individual training system, the responsibility of Army schools and units, involves defining jobs, selecting critical tasks to be trained, developing and conducting training, and evaluating performance. The first step in this process involves determining the characteristics of each MOS. According to U.S. Army Training and Doctrine Command (TRADOC) Pamphlet 350-30, Interservice Procedures for Instructional Systems Development (IPISD), job characteristics can be determined through a variety of methods, including on-site interviews and observation, expert juries, group interviews, and survey questionnaires. Army schools indicate that they do follow the guidelines set forth in IPISD but an assessment of the extent of their compliance is beyond the scope of this study. However, it can be noted that a potential complication in complying with these regulations stems from the present methods used for gathering job analysis data. The job analysis data used by TRADOC for Army MOSs is supplied by the Army Occupational Survey Branch (AOSB), which is part of the U.S. Army Military Personnel Center. AOSB obtains its job analysis data by administering occupational surveys to job incumbents and analyzing the resulting data with the Comprehensive Occupational Data Analysis Programs (CODAP), (Christal, 1974).

Job analysis is a central element in the overall training development process because it is the primary basis for the development of Soldier's Manuals and the identification of critical tasks for training. In addition, the job analysis may be used in the development of enlistment predictors and job performance measures. In order for the products of a job analysis to be effectively used in the training system, the analysis should have certain features:

- o the survey results should contain complete and accurate information from participants
- o surveys should have a reasonably short development cycle so that users will receive products relevant to the current state of jobs
- o surveys should be provided for all MOSs that are part of the system
- o standard procedures should exist for using job analysis products

The ability of current AOSB products to meet each of these standards can be questioned. Job analysis data are obtained by asking job incumbents to indicate the activities they perform on a survey consisting of detailed task and equipment lists, often exceeding 700 items. Such lengthy surveys require extraordinary diligence on the part of job incumbents if accurate, complete data are to be obtained. Furthermore, these surveys are completed and returned on a voluntary basis. (The AOSB reports a 70 percent return rate.) These two factors, lengthy surveys and

voluntary completion, may act to reduce the representativeness and reliability of the resulting data.

Secondly, the AOSB reports the developmental, administrative, and analytic time for a survey to be between 12 and 15 months. During this time, a job can change significantly. The problem of jobs changing after the occupational survey is completed is compounded by the fact that AOSB does not conduct a survey on each MOS in each development cycle. Thus, it is not uncommon for seven years to elapse between surveys for a given MOS.

Finally, Army schools, the most obvious user of AOSB products, have the prerogative of deciding whether to use AOSB products, and if they do so, how to use them. While TRADOC Pamphlet 350-30 does contain guidance on the use of job analysis products, there does not appear to be uniform application of it.

ENTRY-LEVEL TRAINING

Schools which provide entry-level training for an occupational specialty have an effect on job performance assessment insofar as they train soldiers to perform some of the tasks upon which they will be evaluated on the job (e.g., on a SQT). Schools are also responsible for providing support for training within units. A Soldier's Manual contains many more tasks than can be trained during entry-level training. Therefore, schools must decide which tasks or skills will be trained in schools and which in units. If a school is constrained by a lack of equipment or other resources, tasks that otherwise might be trained at a school will have to be trained after soldiers are assigned to field units that have the equipment.

The selection of tasks for entry-level training may be made on a variety of different criteria. For example, schools may choose to train:

- The most difficult or complex tasks.
- Tasks which a soldier is most likely to perform in his first assignment. (These are likely to be simple, rather than complex tasks.)
- Generic skills that would apply across a variety of equipment (especially likely in cases where a school does not have access to all the equipment in the field or when equipment has a short life-span).
- Tasks on which soldiers do poorly on the SQTs. This raises the question of whether poor performance is due to inadequate training in school, inadequate training on the job, or invalid or unreliable task tests.

In order to make these decisions, schools use guidance provided in TRADOC Pamphlet 350-30 but they also require additional information concerning:

- The length of time it takes for an average soldier to master a particular task in the field. This information is not currently provided.
- The likelihood of soldiers performing various tasks once assigned to a unit. These data are provided by AOSB.

- The life-span of particular equipment, technical procedures, official forms, or regulations which a soldier is trained to use. This information is generally accessible to the schools.

- The tasks which soldiers in the field have difficulty performing, as determined by performance on SQT or other means. These data are provided by the Army Training Support Center and/or by units in the field.

The extent to which schools receive and use these types of information will influence the effectiveness of the entry-level training they provide.

SKILL QUALIFICATION TESTS (SQTS)

The purpose of this section is to assess the effectiveness of the SQT, as it is currently composed, as a criterion measure of individual soldier performance. This analysis must begin with the caveat that the primary purpose of SQT is not to serve as a measure of individual performance of soldiers in their trained MOSs. Rather, its intended purpose, as noted in TRADOC regulation 351-2, is as a training diagnostic. As such, an SQT may have properties that are inconsistent with desired features of a criterion measure. Therefore, some of the criticism of SQTs presented in this report addresses characteristics of the test it was not designed to possess. The question being asked here is to what extent the SQTs, as currently constituted, could serve as valid and reliable criteria of individual soldier performance.

SQT Design

To assess the adequacy of the design of SQTs as a criterion measure, the desirable features of such a measure should be identified. Appendix K of this report, Assessment of Job Performance Measures -- System Requirements, contains such an analysis. Psychometric considerations such as validity, reliability, and fidelity must be examined as well as more practical considerations such as face validity, user acceptability, and test development and administrative costs.

Job-Site Component (JSC)

The JSC is administered by a direct supervisor. The procedure involves observing a soldier performing specific tasks in an actual job setting. This represents an attempt to achieve a simulation highly representative of the job. The dilemma of simulation, however, is that increased fidelity, while appearing to be associated with increased validity, may also be accompanied by decreased control and thus decreased reliability (Fitzpatrick & Morrison, 1971). For example, suppose a supervisor of Personnel Administration Specialists (75B) were to observe several skill level 1 soldiers preparing request forms for personnel action (DA Form 4187). He might decide to measure the speed and accuracy with which soldiers completed these forms. However, he would have little control over the situations requiring these forms to be completed. A request for a change of name due to marriage may be a less difficult action than dropping a soldier from the rolls as a deserter. The complexity of regulations for these actions, demands for accompanying documentation, and number of entries required may vary for these two situations, precluding

fair comparison of speed or accuracy scores. In the example given above, the test conditions are highly representative of the job but evaluation is highly complex.

In addition, a supervisor is responsible for training his subordinates and insuring their competence. Furthermore, supervisory judgments on the JSC (to which the subordinate has access) can be expected to affect soldier morale. As a result of these factors, it is difficult for a supervisor to be accurate and objective in his judgments.

The extraordinarily high pass rate on JSC tasks, which exceeds 98 percent for the specialties sampled in this study, indicates that this portion of the SQT cannot discriminate between different degrees of competence. If the JSC measures the same types of competencies as the Hands-On and Skill Components, the fact that pass rates are lower on the Hands-On Component and much lower on the Skill Component suggests that JSC measures may be inflated. This interpretation is consistent with the low correlations between JSC scores and AFQT/ASVAB scores for the eight Army MOSs in this study (See Section II of this report).

Hands-On Component (HOC)

The HOC requires soldiers to perform job-relevant tasks under highly standardized conditions, including a formal test site, trained scorers, and actual equipment or simulators. This procedure represents an attempt to simulate the job while still retaining control over the testing situation.

The HOC appears to have great potential as a criterion measure, but its potential has not been realized due to the way it has been administered in most Army units. The problem stems from the fact that routinely soldiers are run through a practice HOC a week or less before the official test. This practice run is usually identical in every respect to the operational administration of the test. Indeed, at one installation, two of these practice sessions are mandated. While this procedure is laudable in that intensive training is provided on the selected tasks, it tends to contaminate the scores as criterion measures.

First, by offering practice on the tasks to be tested, performance on the HOC overestimates the baseline level of performance of soldiers on these tasks. Second, because the training stimulated by these practice sessions is limited to those tasks to be tested, one cannot infer from HOC performance the level of competence soldiers possess for the untested task domain.

As a result of those characteristics described above, the level of performance on the HOC for the eight Army MOSs averages better than 80 percent. The correlation between this measure and the AFQT/ASVAB is fairly modest (i.e., less than .30). Another reason for the modest relationship between HOC and AFQT/ASVAB scores is that performance on the HOC is scored on a task level as Go/No Go. Each task has a number of items (SC) or performance measures (JSC and HOC) and the cutoff score for each task test is set individually. Examinees who pass a task test may have performed acceptably on varying numbers of performance measures. Thus, when these task test results are categorized as Go/No Go, information is lost. The preservation of test results at the item and performance measure level may therefore enhance the correlations between HOC and AFQT/ASVAB scores.

Skill Component (SC)

The SC is a performance-based, paper-and-pencil test which checks the ability of a soldier to perform certain tasks or to apply the knowledge necessary to perform a task. On the SC, a soldier is asked to read and answer a set of written, multiple-choice questions.

The SC test has the potential to be a cost-effective measure of job performance. In order to realize its potential, however, close attention must be paid to several features of the test.

Task selection for the SC (as well as for the other components of the SQT) is made formally in the SQT plan 15 months prior to the start of the test period. The job and task analysis which are used to develop the test plan may have been conducted considerably earlier. As a result, especially for MOSs in which equipment or regulations change frequently, test items can become obsolete before they are fielded.

It is generally agreed that writing effective performance-based multiple-choice questions requires significant levels of subject matter expertise and test writing skill. Therefore, content experts and test development specialists are mandated for writing SQTs. However, as we learned from our field interviews, in many cases only content experts are provided. Senior NCOs may be assigned, regardless of their career preference, to serve as content experts. They usually have no prior training or experience in test writing. They typically attend a two-week training course or are provided with workshop materials on test writing.

To assess the current state of SC test items produced by content experts, a technical evaluation of a sample of 1980 SC test items was conducted for this report. All eight MOSs in this study were included in the assessment. The criteria used for evaluation (c.f., TRADOC Pamphlet 351-2, pages 94-97; Wesman, 1971; Thorndike and Hagen, 1977) included whether:

- o questions were written in clear, simple language
- o questions avoided trivia, and were related to task performance
- o response choices were parallel and realistic.

The assessment indicated that violations of one or more of the criteria were present in over 30 percent of the items sampled. This finding is consistent with the general observation that item writing skill requires several months to develop.

Another approach toward assessing the adequacy of the training provided for test writers is to examine the materials used in training them to perform a particular step in the test-writing process. During our field visits, we questioned test writers about their training and reviewed the guidance and materials available to them. For example, component task analysis is the process by which a test writer identifies the critical elements of a task so that they can be addressed with specific test items. This process receives only brief attention in the current training curriculum for test developers; examples are provided from only a limited number of specific skill areas (e.g., mechanical maintenance). The failure to tailor such

examples to the particular skills and knowledges of prospective test writers and the brief treatment the topic receives, combine to insure that the ability to do task analysis will be insufficiently developed. Another skill which seems to receive inadequate coverage in test writing training is the development of test item distractors and their relationship to performance standards.

After test items for the SC are written, they are validated according to procedures detailed in TRADOC Pamphlet 351-2. Test items are administered to soldiers in the field to determine whether the items discriminate between performers and nonperformers. This approach to validation requires both the accurate identification of performers and nonperformers independent of the test items themselves (external validation) and the selection of test items which discriminate between performers and nonperformers.

The optimal method for identifying performers and nonperformers, a hands-on performance rating method, is resource intensive and has been judged infeasible. The method most commonly employed is a self-rating method. This method involves having soldiers read descriptions in Soldier's Manuals of designated tasks and asking them if they can perform these tasks to required standards. Soldiers claiming they can perform a task are called "performers" while the remaining soldiers are called "nonperformers" for that task.

Subsequently, both "performers" and "nonperformers" take the SC task test. Items which produce agreement with the self-rating classifications (that is, items on which "performers" score as high or higher than "nonperformers") are considered valid. The main problem with this method is that there is little confidence that soldiers can accurately classify themselves as performers or nonperformers. In fact, in the case of the one MOS for which we had data on all validated tasks, "performers" scored 48 percent correct and "nonperformers" scored 40 percent correct, an unimpressive margin. The lack of apparent discrimination between performers and nonperformers using the self-rating method and the fact that the minimum criteria for item selection do not require each item to significantly discriminate performers from nonperformers combine to raise questions about the effectiveness of existing validation procedures.

In the absence of an acceptable index for identifying performers, other approaches to item selection should be considered. The original conception of criterion-based measurement specified that test items were to constitute a sample of behavior from a well-defined behavioral domain. Exhibited performance levels would be "directly interpretable in terms of a performance standard" (Glaser & Nitko, 1971). The use of conventional item analysis techniques for selecting final test items was eschewed because the application of such techniques leads to the selection of a set of items that are not representative of the original behavior domain (Davis & Diamond, 1974).

Such a "content validity" approach to item selection requires not only careful task and job analysis, but also a systematic approach for generating items from behavioral descriptions (e.g., Hively's "item forms") in order to allow for estimation of "true domain scores." Current SQT development procedures do not meet this level of precision and, indeed, probably could not do so within the constraints imposed by the breadth of the job performance to be measured and the time available for test development and for test administration.

Alternative item selection procedures developed for criterion-referenced measurement involve selecting items that show sensitivity to instruction. Items are selected for the final test only if there is a sizable difference between the proportion of examinees passing the item before and after instruction (see Mehrens & Lehmann, 1978). Such a procedure would be reasonably cost-effective to implement and would insure that training and testing cover the same skills. (However, the relationship of those skills to job performance is not established empirically by this method and hence still depends upon the integrity of the job and task analysis.)

Performance is lower on the SC than on the other two components of the SQT. The average soldier in the eight MOSs included in this study received a Go on less than 50 percent of the tasks tested in the SC. The corrected correlations between SC scores and AFQT scores were impressive, ranging between .40 and .55 for the MOSs in this study. Further, the corrected correlations of SC scores with aptitude composite scores were even higher, ranging from .49 to .63. However, these correlations do not, by themselves, provide conclusive evidence of the validity of the SC. The SC must be demonstrated to be a valid measure of job performance before these correlations can be accepted as evidence that the ASVAB is a valid predictor of success in various occupational specialties.

The problem of specifying an independent performance criterion is equally relevant to test validation. It is unlikely that any single, "ultimate" index of job performance can be identified. However, various measures of job performance on the same or related tasks should correlate, provided those measures yield scores with sufficient variability. If the HOC and JSC can be modified to make them more discriminating, the demonstration of a relationship with the SC would validate that component of the SQT as a measure of job performance.

Finally, the 'cutoff' or passing score on the SQT has been arbitrarily set at 60 percent. This standard is admittedly arbitrary, and does not take into consideration the differences in difficulty or importance of the tasks included within an SQT or on the different SQTs employed in different specialties.

A widely used technique for establishing performance standards is to seek the judgment of a panel of experts in the relevant field. For each task to be tested, subject matter experts could be asked to judge the minimum number of items a competent soldier could be expected to pass. Such approaches have been criticized as arbitary (Glass, 1980). However, they are less arbitrary than the current practice and, as Hambleton (1978) argues, at least reflect professional judgment in the field.

Ideally, the SQT cutoff score would be based on an empirically established relationship to job performance. However, this undertaking again depends upon an elusive, validated measure of job performance. A composite measure based on multiple individual measures is likely to be the best feasible alternative for standard setting as well as for validation and item selection.

Summary

Each component of the SQT has deficiencies that serve to limit the SQT's effectiveness as a criterion measure. In addition, since any SQT measures only a small number of the tasks which comprise a soldier's job, it is important to be able to infer from test performance a comparable level of competence for the tasks not included on the test. However, in the case of the SQTs, participants are notified as to which tasks will be assessed 90 days prior to SQT administration. Indeed, in most cases detailed descriptions of how each task will be tested are provided. As a result, inferences regarding the competence of soldiers on tasks not in the test domain cannot be made.

SECTION V

ASSESSMENT OF TRAINING PERFORMANCE MEASURES

INTRODUCTION

Schools which provide entry-level training in an MOS serve a number of needs. First, schools train soldiers to perform some of the tasks which they will be required to perform on the job. These tasks comprise the Program of Instruction (POI) for a training course. Some of the considerations in developing a POI as well as a brief review of process are provided in Section IV (Entry-Level Training).

Second, schools provide criteria of performance in training. These criteria may serve a number of purposes, including:

- o to make decisions regarding the acceptability/unacceptability of performance in training (e.g., to pass a "gate" test or to complete training with an acceptable final course grade). Those who succeed in training should succeed on the job.

- o to validate prediction tests (e.g., AFQT/ASVAB). AFQT/ASVAB is used to determine eligibility for entry into various occupational specialties and therefore must be able to differentiate among various levels of performance in training and on the job.

A third function of schools is to provide support for training in units. This support can include instructional materials for tasks not trained during entry-level training, job performance aids, etc. The analysis conducted in this study focused on the adequacy of training criteria.

First, in order to assess the relevance of training and training criteria to the job, POIs and course tests were compared to Soldier's Manuals. The adequacy of test administration and scoring procedures was evaluated by interviewing test administrators and by observing test sessions. Test development procedures were assessed by interviewing test developers, reviewing official test development guidance and test development training materials, and by conducting a technical evaluation of course tests. Test validation procedures were ascertained through interviews of test development personnel. The adequacy of existing and alternative training criteria for validating predictors was evaluated by first obtaining samples of soldiers in the entry-level training course for each occupational specialty. Criterion measures were obtained for each of these samples and related to AFQT/ASVAB scores for the same individuals.

Relevance of Training and Training Criteria to Job

In order to insure the connection of training and training criteria to job requirements, the following conditions must be met.

- o POIs must contain key performance elements of jobs. This subject is addressed in Section IV of this report under the heading, Entry-Level Training.

- o Training and training criteria must reflect the material in the POI. Soldier's Manuals, POIs, and course tests for the MOSs studied were compared in terms of content, and a strong correspondence was found.

Standardization of Test Administration and Scoring

The administration of written and/or performance tests was observed for all initial entry training courses. Interviews were conducted to determine test scoring methods, procedures for ensuring test security, etc. The findings were:

- o test administration of written and performance tests was highly standardized, scoring methods seem to be reliable, and test security was generally adequate (although test security could be improved if alternate forms were developed or if item pools were established)

- o in lockstep training courses, the small number of test stations for performance tests resulted in trainees having to spend a lot of time either waiting to be tested or waiting for others to finish their tests.

Test Development

Procedures for developing tests are provided in TRADOC Pamphlet 350-30. Tests found in training courses were generally of two types, written multiple-choice tests and performance tests.

Writing effective multiple-choice tests, as indicated in Appendix G of this report, requires significant subject matter expertise as well as test writing skill. To assess the state of multiple-choice test items produced for training course tests, samples of test items were evaluated for technical adequacy. The assessment indicated a number of problems including:

- o answer choices with "length" cues

- o "all of the above" answer choices

- o ungrammatical distractors

- o the use of low-level knowledge questions to test higher level concepts

- o unrealistic distractors.

Further, the questions on a number of the tests appeared to be unordered. This was especially true in the case of computer generated tests. In computer generated tests having supplements containing figures, it was quite a difficult task to keep track of appropriate figures for each test question.

Another issue of significance to test developers concerns the reading level of test materials. These materials should require reading skills no higher than do

materials required for the job. Little evidence was found in the schools that this issue is directly addressed.

The development of written and performance tests share another problem; that is, insufficient training of test developers to identify the critical elements of tasks so that they may be specifically addressed by test items or performance measures.

Test Validation

A number of issues related to the validation of tests and the use of test results will be addressed here including:

- o tryout of tests, predictive validation
- o item selection, identification of problems
- o setting of performance standards
- o reliability.

Based on our field interviews of test development personnel, we learned that in most entry-level training courses, tests do not receive tryouts before implementation, nor are predictive validation studies conducted. Thus, it is unclear how items are selected for inclusion on such tests, whether selected items discriminate performers from nonperformers, and how performance on these tests relates to competence.

While little evidence was obtained concerning procedures for selecting items before test implementation, once tests were implemented some schools periodically examined the difficulty values (p, proportion of test takers who answered a test question correctly) of items on tests. Items whose p values were less than a specified value (usually 80 percent) were reviewed to determine whether poor performance was attributable to failures in instruction or to a poorly written test item. While this practice, by itself, is exemplary, it excuses easy test items from such a detailed review.

Performance standards on tests in training courses were set arbitrarily. No evidence was available linking levels of performance in training to independent measures of performance or to subsequent performance on the job.

Estimates of test reliability were generally not available for school tests. However, most written tests contained a sufficient number of items (i.e., 50 items) to insure a minimum level of reliability. (This is true only if the tests are internally consistent and measure the same dimension.)

Test Administration

In general, test administration procedures were observed to be highly standardized. Test compromise appears to be unlikely considering the careful attention given to the accounting of test booklets and answer keys. In some specialties, the

repeated use of the same test booklets, which picked up some marks, could be somewhat of a problem.

FINAL COURSE GRADES

The current final course grade measures have a number of limitations. With regard to the ability of such measures to make distinctions regarding the acceptability/unacceptability of performance in training and subsequently, on the job:

- unless job analysis products, and therefore training curricula, accurately reflect the jobs for which trainees are being prepared, job measures cannot reasonably be expected to distinguish between performers and nonperformers.

- the validity of training measures remains undetermined because validity and reliability studies of criteria were not conducted. In addition, questions can be raised concerning the qualifications of test developers.

- final course grades in training courses organized in a self-paced manner do not present a complete picture of a trainee's performance, but should be considered together with time-to-complete factors.

- the failure to link levels of performance on training criteria to levels of subsequent performance on the job raises serious doubts about the accuracy of current acceptability/unacceptability decisions.

A second capability of training criteria is their ability to discriminate among various levels of performance. If measures meet this standard, they should have utility for validating predictor tests (e.g., AFQT/ASVAB). The training criteria for the occupational specialties examined in this study have some limitations in this regard, as demonstrated by:

- the fact that final course grades were themselves never validated as criteria.

- the high and somewhat attenuated distribution of final course grades (e.g., means range from 81 to 94 while standard deviations range from about 4 to 10).

Despite these limitations, however, some fairly high correlations were obtained between AFQT/ASVAB scores and final course grades (i.e., as high as .68).

ALTERNATIVE EXISTING TRAINING PERFORMANCE MEASURES

This class of performance measures consists of additional information that is currently available which is related to training performance. Two of these measures --attrition and time to complete training--depend on, in part, performance on measures which are used to calculate final course grades.

Attrition is a dichotomous measure. The decision leading to dropping an individual from a training course is based largely on a comparison of obtained grades with cutoff scores. As stated earlier, the fact that these cutoff scores were arbitrarily designated does not give confidence in decisions based on these scores. In addition, dichotomous measures like attrition are difficult to use for validating predictors since they cannot distinguish between more than two levels of perfor mance. New statistical techniques such as maximum likelihood estimation allow more precision in predicting dichotomous criteria (Dempsey, Sellman, & Fast, 1979).

Time to complete training is based on the rate at which an individual proceeds through a training course. Such courses are usually marked by a series of "gate" tests which an individual must pass in sequence in order to graduate. Thus, time-to-complete is linked to existing training performance tests. However, time-to-complete can also be interpreted as a rather direct measure of relative cost to train in school and, possibly, on the job. It does have some limitations, however. For example, if soldiers are not motivated to complete training as quickly as possible, then time-to-complete may reflect motivation as well as ability.

Finally, alternative performance tests like the Mortar Qualification (MQ) test may have some value as criteria. In MOS 11C (Indirect Fire Infantryman), final course grades average 93 (SD=4.78) while the MQ test scores average 74 (SD=12.49). The MQ measure is less attenuated than final course grades and is more highly correlated with AFQT/ASVAB scores than final course grades. Thus, measures like the MQ test may make a contribution to training performance measurement.

SECTION VI

ASSESSMENT OF ALTERNATIVE JOB PERFORMANCE MEASURES

This section of the report contains a description of the alternative job performance measures developed for this project. The tryout of these measures is discussed and the potential for such measures is assessed.

RATIONALE FOR ALTERNATIVE JOB PERFORMANCE MEASURES

The alternative job performance measures developed in this project were not designed as a replacement for the SQT and Enlisted Evaluation Report (EER) instruments used currently. Within the scope of this study, an instrument matching the three-part SQT in comprehensiveness was not feasible. The purpose in designing an alternative measure here was to find a cost-effective instrument that would fill in some gaps left by present measures.

The SQT includes a written test of job knowledge (SC), a hands-on performance test (HOC), and a set of supervisor ratings of actual job performance (JSC) as described in Section II. Of these three types of measures, supervisor ratings are the easiest and least expensive to collect. However, there are weaknesses in the JSC ratings (as discussed in Section IV), and the alternative measures developed for this project were designed to yield supervisor ratings that would not suffer from the same limitations.

The JSC asks for task performance ratings for an individual soldier on selected MOS tasks even though the soldier may not perform some of those tasks at his current job, or his supervisor may not have observed him performing them. Such a situation is bound to decrease the validity and reliability of these ratings. Moreover, the JSC takes into account neither individual differences in conditions under which tasks are performed nor differences in the difficulty of the tasks comprising various soldiers' jobs. Finally, since soldiers are rated simply on a Go/No Go basis, the JSC is not very good at differentiating among different levels of competence and has a severe ceiling problem (98% Go decisions). If nearly every soldier is going to be rated as competent on every task, there is little point in administering the instrument.

The other supervisor rating form used at present, the EER, uses a five-point scale but is not task-based. Instead, supervisors are asked to rate soldiers on a series of very general dimensions (e.g., adapts to changes, attains results, integrity, moral courage, earns respect).

The goal in developing new measures was to obtain a reliable, sensitive set of supervisor ratings on the tasks individual soldiers actually perform at their respective jobs, taking into account the number of tasks performed and their difficulty. These ratings should discriminate among various competency levels well enough to be used as a criterion for evaluating predictor measures.

DESCRIPTION OF NEW MEASURES

The goals delineated for alternative measures of job performance necessitated determining

- o which MOS tasks each soldier performs
- o the difficulty of each task
- o how well each task is performed.

Three instruments were developed to fill these functions. The Occupational Survey consists of a list of all MOS tasks at the appropriate skill level grouped according to functional or equipment characteristics. Each soldier is asked simply to check off each task he or she currently performs on the job.

The Task Difficulty Rating Form, which is completed by supervisors, contains the same list of MOS tasks, and requires the supervisor to rate, on a five point scale, the time it takes to learn to perform each task.

A Job Performance Rating Form is filled out for each soldier by his or her direct supervisor and calls for performance ratings on each MOS task the soldier currently performs. Ratings are made on a five-point scale with three being "meets required standard". This instrument yields a list of all the MOS tasks each soldier's supervisor believes that soldier currently performs, a numerical estimate of the soldier's level of performance on each task, and a mean performance rating for the soldier (averaged over the tasks that soldier performs).

Copies of the three experimental job performance measurement instruments developed for MOSs 11B (Infantryman), 31M (Multichannel Communications Operator) and 75B (Personnel Administration Specialist) are contained in Appendix L. Only the first two pages of each form are included.

In addition to the data obtained from each individual instrument, useful measures can be derived by combining data from several instruments. For example, average difficulty ratings can be computed for the tasks performed by individual soldiers and composite job performance measures can be developed which take into account a soldier's performance level on individual tasks, the difficulty of those tasks, and the number of tasks the individual performs.

TRYOUT PROCEDURES

During visits to Fort Bragg and Fort Hood in July 1981, the new instruments were administered to 59 soldiers and 24 supervisors in MOS 11B (Infantryman), 50 soldiers and 27 supervisors in MOS 31M (Multichannel Communications Operator), and 19 soldiers and 12 supervisors in MOS 75B (Personnel Administration Specialist).

Soldiers and supervisors responded to the surveys separately, making it possible to compare soldier and supervisor reports of which tasks each soldier performs. Instructions for the surveys included an acknowledgment that an

individual soldier probably would not perform all the tasks for his or her MOS and instructions that ratings should be provided or tasks should be checked only for tasks the soldier was performing in his or her present job. Respondents were informed that results from the survey would not be used to evaluate an individual or a unit.

FINDINGS

The three instruments employed in this project were designed as complementary components in a system for measuring each individual's job performance. Information on which tasks each soldier currently performs (gathered from the Occupational Survey and from the Job Performance Rating Form) and on how well each of those tasks is performed (in the supervisor's opinion) can be weighted according to the number and difficulty of those tasks. Soldiers performing more tasks or more difficult tasks contribute more than soldiers performing fewer/easier tasks at the same level of proficiency.

A. Occupational Survey

The first step in individual job performance measurement is the determination of which tasks each soldier currently performs. The SQT program assesses all soldiers on a sample of tasks from their MOS, regardless of whether or not those tasks are ones the soldier performs on the job.

The particular tasks performed by each soldier in the present study were ascertained in two ways - by asking the soldier in the Occupational Survey and by asking the soldier's supervisor as part of the Job Performance Rating Form.

Two sources of data were used here because earlier research (Sellman, 1968) suggested that soldiers may provide more accurate information concerning which tasks they perform than do their supervisors. Since supervisors were going to be providing proficiency ratings on those tasks each soldier currently performs, supervisors would necessarily be making judgments on which tasks the soldier performs at the same time they rated how well the soldier performs them. A very high rate of agreement between soldiers and supervisors concerning which tasks the soldier performs would suggest that future surveys could be administered to supervisors only.

Such agreement was not found for the MOSs in this study, however. Adding the first and last columns in Table 18 reveals that the percentage of tasks on which a soldier and his or her supervisor gave the same report ranged from 74% in MOS 11B to 90% in MOS 31M. These indices are inflated, however, by the inclusion of large numbers of tasks on the surveys which both soldiers and supervisors agree the soldiers do not perform (45% of the tasks in MOS 11B, 74% in 31M, and 50% in 75B). If agreement is figured only for tasks which the soldier, the supervisor, or both indicate are performed, the two types of reports agree for only 53% of the soldier/task combinations in MOS 11B, 62% in 31M, and 52% in 75B (from Table 18: (a)÷(a)+(b)+(c)).

Table 18

Agreement Between Soldiers and Their Supervisors on Tasks Performed

Percentage of MOS Tasks

MOS	Soldiers & Supervisors Report Performed (a)	Soldier Only Reports Performed (b)	Supervisor Only Reports Performed (c)	Soldiers & Supervisors Report Not Performed (d)
11B	.28	.16	.09	.45
31M	.16	.03	.07	.74
75B	.26	.08	.16	.50
Overall	.25	.13	.09	.53

Implications

The surprisingly low level of agreement concerning which tasks each soldier performs suggests that attempts to provide a job profile for each individual soldier should not be based solely on supervisor reports.

Further, these data and the low percentage of MOS tasks which both soldiers and their supervisors agree the soldier performs (28% for MOS 11B, 16% for 31M, 26% for 75B) imply that there may be inadequacies in the communication of which tasks soldiers are to perform. Occupational data of the sort gathered in this project are inexpensive to collect and analyze, and have potential as a tool in detecting possible weaknesses at the individual or the unit level.

We recognize that lack of agreement in these reports may stem from a variety of different sources:

- Soldiers and supervisors may differ in their interpretation of task statements. Supervisors may be more familiar with the language used in the Soldier's Manual; soldiers may not recognize tasks which they actually perform from the verbal description on the survey form. This possible artifact would contribute to the percentage of tasks supervisors indicate a soldier performs that the soldier himself says he does not perform (7% to 16% of the task/soldier combinations).

- Soldiers and supervisors may differ in their interpretation of survey instructions. Although both groups were explicitly told to indicate only those tasks that each soldier performs at his or her current job, soldiers and supervisors may differ in the way they interpret these instructions. In addition, the two groups may, to different degrees, feel reluctant to admit that soldiers do not perform certain MOS tasks.

- Supervisors may not know what their supervisees are doing. This condition may result because Soldier's Manuals are unclear, Soldier's Manuals are not being used, or because personnel receive inadequate supervision.

Data of the sort gathered in this project cannot reveal the source of particular areas of disagreement, but they can be used to locate either units or individuals where very low levels of agreement warrant further investigation.

B. Difficulty Ratings

Mean difficulty ratings were obtained for each Soldier's Manual task for the three MOSs in this study. These values range from 1.46 to 4.07 for MOS 11B, from 2.00 to 4.13 for 31M, and from 1.58 to 3.92 for 75B. Thus, supervisors did discriminate among tasks of different difficulty levels (SD=.67 for MOS 11B, SD=.56 for 31M, SD=.59 for 75B). Average difficulty ratings (2.69 for MOS 11B, 2.75 for 31M, and 2.69 for 75B) were slightly below the mid-point for the five-point scale.

C. Job Performance Ratings

Performance ratings for those jobs an individual soldier currently performs provide both an alternate measure to the SQT and EER and useful information for planning training.

A major problem with JSC and EER ratings has been the lack of variability. With nearly all soldiers receiving positive ratings on a binary variable as on JSC Go/No Go ratings, there is little discrimination among levels of competence and insufficient range for the ratings to serve as criteria against which predictive measures can be validated. Ratings on abstract general scales (e.g., "moral courage") of the sort used on the EER are not directly related to task performance and often prove unreliable.

One of the major innovations in these alternate measures was the use of a five-point scale to evaluate performance on specific tasks. This modification was designed to increase the variance in soldier ratings and to eliminate the ceiling problem in the ratings. This effort was successful. Mean job performance ratings ranged from 1.43 to 4.81 with a mean of 3.29 and a standard deviation of .94 for MOS 31M, from 1.00 to 4.45 with a mean of 3.08 and SD of .82 for 11B, and from 1.00 to 5.00 with a mean of 3.11 and an SD of 1.34 for 75B.

These data are in marked contrast to the ratings typically received on the JSC as shown in Table 19 on the following page. While the JSC as currently administered locates few cases of below-standard performance, the measure employed in this field test found performance to be inadequate around 40% of the time.

It must be recognized that this comparison pits an experimentally administered new measure against data derived from an operationally administered measure. Several of the factors reducing the variance in JSC scores would also affect an operational administration of the alternative measure. Supervisors who feel that low job performance ratings reflect poorly on them or undermine soldier morale are likely to give inflated competence ratings with any evaluation instrument. However, it may be easier to give a soldier a two on a five-point scale than to fail him or her on a Go/No Go decision. At any rate, the five-point scale provides

adequate range for discriminating among competency levels, and the try-out data are certainly encouraging enough to suggest further investigation.

If the alternative measure is operationalized, ratings can still be used for making Go/No Go decisions by simply converting scores of three or better to Go and those below three to No Go. The difference from current procedures is that enough data would be available to use the alternative measure in validating predictors, and the ceiling problem (98% Go decisions) should be ameliorated through use of a more sensitive measure.

Table 19

Average Rate of GOs on Tasks for
JSC and Alternate Job Performance Rating Methods

MOS	JSC	Alternate Measure
75B	.98	.58
11B	.98	.61
31M	.98	.58

SECTION VII

SUMMARY AND RECOMMENDATIONS

SUMMARY

This section provides a brief review of the principal findings of this study. This review is divided into two parts, the first dealing with job performance measures and the second dealing with training performance measures.

Job Performance Measures

The findings from the assessment of the Skill Qualification Test (Section IV) are summarized below:

- SQT scores are positively related to AFQT and to aptitude composite scores; that is, recruits with high AFQT/ASVAB scores also score high on the SQT. Therefore, AFQT does predict SQT performance.

- While the strength of the relationship between AFQT/ASVAB and SQT scores is about the same for whites, blacks, and Hispanics, AFQT/ASVAB scores tend to overpredict the SQT performance of blacks (and to a lesser extent Hispanics).

- The SQT is positively related to graduation from high school. High school graduates perform better than non-high school graduates (with equivalent AFQT/ASVAB scores) on the SQT.

- An examination of the relationship between AFQT/ASVAB scores and each of the three components of SQT should take into account that the Skill Component (SC) is a written test like the AFQT/ASVAB and that Skill Component test scores have much more variance than HOC or JSC scores. Thus, the finding that AFQT/ASVAB correlates more highly with the Skill Component than with the Hands-On (HOC) or Job-Site Component (JSC) is not surprising.

- The Skill Component, a written test, can simulate job performance most readily for skill levels/jobs of MOSs in which cognitive skills are important (e.g., administrative/technical M[illegible] nanagerial positions). Alternatively, the Hands-On and Job-Sit[illegible] onents can simulate job performance most readily for the MOSs in which psychomotor skills are important (e.g., skill level 1 combat arms MOSs). Thus, the SQT system must be, but presently is not, flexible enough to accommodate the inherent differences among various MOSs.

- Average scores on the SQT increased by an average of 20 percent from the first to the second year of fielding. Such an increase may reflect a stronger command emphasis on unit training, or indicate that the tests are partially compromised in the process of field administration.

- Conditions which limit the reliability, validity, and/or potential utility of the SQT as a criterion are as follows:

  - The current self-rating method of identifying performers and nonperformers during the field test of SQT items is ineffective. It leads to the selection of test items in many cases which do not adequately differentiate between performers and nonperformers.

  - The number of items used to test a particular task in the SC is not currently based on considerations of reliability or task criticality. Rather, it is merely determined by the number of items developed which meet the criteria for inclusion on the test.

  - The technical quality of the SQT is lowered when test items are written by personnel who lack sufficient test writing knowledge and skills.

  - SC scores may be affected by examinees' reading ability even in specialties and skill levels where strong reading skills are not required (e.g., skill level 1 of many combat arms specialties).

  - Tasks tested on the HOC and JSC are unit weighted and scored on a Go/No Go basis. Task performance is not weighted according to importance (criticality) or difficulty. Moreover, rather than receiving a numerical score based on individual performance measures, examinees receive a Go if they perform acceptably on a task, or No Go if they perform unacceptably. These features contribute to the finding that HOC and JSC tests do not adequately distinguish between various levels of performance.

  - The Hands-On Component is usually practiced within units 1-2 weeks prior to the actual test. In addition, SQT Notices, which are released 90 days prior to test administration, specify exactly which tasks will be tested on all components of SQT and how they will be tested. These practices seriously compromise the test results. As a result, the HOC and to a lesser extent the SC and JSC, rather than measuring typical performance, measure maximal performance. Furthermore, training emphasis is placed only on those tasks tested on the SQT.

  - Supervisors are lenient in their ratings of subordinates on the Job-Site Component. It is difficult for them to criticize soldiers whose performance they are held responsible for. This leniency results in extraordinarily high pass rates on the JSC (typically exceeding 98%) reducing it's ability to discriminate between performers and nonperformers.

  - SQT test items may become obsolete in the 12-16 months between the time the test is initially developed and when it is administered, especially with the introduction of new equipment or changes in doctrine.

In addition to assessing the SQT, experimental job performance measures were developed as part of this study (i.e., Occupational Surveys for both soldiers and their supervisors, Task Difficulty Rating Forms, and Job Performance Rating Forms). It was found that:

- o There was substantial disagreement between soldiers and their supervisors regarding the tasks performed by the individual soldiers.
- o Soldiers perform tasks that vary substantially in degree of difficulty. There is general agreement among supervisors concerning the relative difficulty of specified tasks.
- o Job performance ratings were variable enough to distinguish among various levels of performance.

Training Performance Measures

The findings from the assessment of training performance measures (Section V) may be summarized as follows:

- o Current training criteria (i.e., final course grades) have a number of weaknesses. First, the distribution of grades is attenuated. Second, the tests have not been validated. Third, cutoff scores have not been validated. Despite these problems, final course grades were positively related to AFQT and to aptitude composite scores.
- o The relationship between AFQT/ASVAB scores and final course grades is stronger for technical and administrative specialties than for combat arms specialties.
- o High school and non-high school graduates (with the same AFQT/ASVAB scores) score at about the same levels on final course grades.
- o Attrition is higher among non-high school graduates than among high school graduates. Attrition is generally not related to AFQT/ASVAB scores. Therefore, attrition may represent a criterion of adaptability to the military.
- o Time-to-complete has potential as a criterion in self-paced courses. Its value, however, may depend on the presence of sufficient incentives for trainees to finish training as quickly as possible. Time-to-complete indices are moderately correlated with AFQT/ASVAB scores.
- o In an experimental setting, peer nomination ratings were moderately correlated with AFQT/ASVAB scores. High school graduates were rated higher than non-high school graduates.

RECOMMENDATIONS

Recommendations are offered in light of the purposes of this study which were to (1) determine the utility of existing training and job performance measures for

validating AFQT/ASVAB and (2) develop experimental alternative training and job performance measures which have potential as criteria for validating AFQT/ASVAB.

The SQT is a valuable criterion of job performance in the Army. The implementation of the SQT has apparently spurred unit training in MOS-specific tasks, resulting in the improved performance of skill level 1 soldiers. While SQT results correlate substantially with AFQT/ASVAB scores, a number of deficiencies in the SQT system were identified, which, if remedied, would enhance the value of SQTs to the Army.

- SQT Notices should contain only a sample of tasks to be tested on the subsequent operational administration of SQT and should not specify, even for these tasks, the exact nature of the test.

- Item selection procedures on the SQT tryout should be based on pre- and post-training discrimination indices or on measures of internal consistency rather than on self-ratings (which is currently the predominant method), at least until better methods are developed (MGA is currently working on a project for the Army Training Support Center to develop more effective procedures).

- Item selection criteria should be changed; difficult items (p values less than .50) should not be automatically excluded from a test, and the criteria for item selection should include a requirement that each item significantly discriminate performers from nonperformers (rather than the current method of simply requiring equal or higher scores from performers than nonperformers).

- Empirical procedures for setting SQT cutoff scores on task tests are effective at linking test performance to performance standards only when performers and nonperformers can be accurately identified. In the absence of such accuracy, subject matter experts should assess the adequacy of task test cutoff scores.

- The practice of having at least two administrations of the HOC, with only the latter administration being operationally scored, should be changed. The first administration should be operationally scored. Subsequent administrations could be used to assess the effects of training.

- The actual scores (number of items correct, number of performance measures Go) on task tests should be retained in calculating total SQT scores.

- The JSC, rather than requiring Go/No Go judgments, should be changed to a multilevel scale (e.g., five points) with behavioral descriptions at each point.

- Training of SQT item writers should be expanded, particularly in the areas of task analysis and technical evaluation of items.

- Greater flexibility shoud be allowed in determining the most appropriate mix of test methods (i.e., SC, HOC, or JSC) for an MOS. Further, consideration might be given to the idea of putting more emphasis on testing specialty-specific tasks in those occupations in which (1) job content remains fairly stable, necessitating less extensive test modifications, and (2) less specialization occurs on the job, making tests more acceptable and relevant to examinees. In general, combat arms specialties meet these requirements to a greater extent than combat support/combat service support. The need for performance measurement in combat support/combat service support specialties might best be satisfied by developing more generic task tests for SQT which (1) will not be sensitive to changes in job content, and (2) will be relevant to examinees who specialize in their jobs.

Several recommendations are also offered with regard to training criteria.

- Serious questions have been raised concerning training criteria (i.e., final course grades), particularly with regard to the lack of validation studies. This research provides an ideal opportunity to conduct such studies. For example, the training samples used in this research could be followed to determine their success on the job. The resulting data would help to determine the predictive validity of existing training criteria as well as alternative and experimental criteria developed in the present study.

- The finding of higher attrition rates for non-high school graduates as compared to high school graduates and the lack of a relationship between attrition and AFQT/ASVAB scores suggests the need to conduct further research to isolate the correlates of attrition.

- Time to complete training could provide a suitable criterion, especially if clear incentives were established for trainees to complete courses as rapidly as possible.

REFERENCES

ASVAB Working Group. History of the Armed Services Vocational Aptitude Battery: 1974-1980. A report to the principal Deputy Assistant Secretary of Defense (Manpower, Reserve Affairs & Logistics). Washington, D.C.: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics), March 1980.

Christal, R.E. (Ed.) Proceedings of 1973 division of military psychology symposium: Collecting, analyzing, and reporting information describing jobs and occupations. (AFHRL-TR-74-19). Lackland AFB, Texas: Air Force Human Resources Laboratory, February 1974.

Dann, J. Selector composite developed from the Armed Services Vocational Aptitude Battery (ASVAB): A tool for predicting attrition from the Basic Electricity and Electronics School. (Special Report 78-4). San Diego, California: Navy Personnel Research and Development Center, February 1978.

Davis, F.B. & Diamond, J.J. The preparation of criterion-referenc i tests. In C. Harris, M. Alkin & J. Popham (Eds.), Problems in criterion-referenced measurement. Los Angeles: University of California, 1974.

Dempsey, J.R., Sellman, W.S. & Fast, J.C. Generalized approach for predicting a dichotomous variable. AFHRL-TR-78-84. Brooks AFB, Texas: Air Force Human Resource Laboratory, February 1979.

Fitzpatrick, R. & Morrison, E.J. Performance and product evaluation. In R.L. Thorndike, (Ed.). Educational Measurement (2nd Ed.), Washington, D.C.: American Council on Education, 1971.

Ghiselli, E.E. The validity of occupational aptitude tests. New York: John Wiley & Sons, 1966.

Ghiselli, E.E. The measurement of occupational aptitude. Berkeley, California: University of California Press, 1955.

Ghiselli, E.E. "Dimensional problems of criteria." Journal of Applied Psychology. 1956, 40, 1-4.

Glaser, R. & Nitko, A.J. Measurement in learning and instruction. In R.L. Thorndike (Ed.), Educational Measurement. Washington, D.C.: American Council on Education, 1971.

Glass, G.V. When educators set standards. In E.L. Baker & E.S. Quellmaly (Eds.), Educational testing and evaluation. Beverly Hills, California: Sage Publications, Inc., 1980.

Goodstadt, B.E. & Yedlin, N.C. A review of state-of-the-art research on military attrition (Final Report). Washington, D.C.: Advanced Research Resources Organization, June 1979.

Greenberg, I.M. Mental standards for enlistment performance of Army personnel related to AFQT/ASVAB scores (MGA Rept. # 0180). Washington, D.C.: McFann, Gray and Associates, Inc., December 1980.

Hambleton, R.K. Test score validity and standard setting methods. In R. Berk (Ed.), Criterion-referenced measurement: The state of the art. Baltimore, Maryland: The John Hopkins Press, 1978.

Hiatt, C.M. & Sims, W.H. Armed Services Vocational Aptitude Battery (ASVAB) and job performance. (CNA 80-3121). Arlington, Virginia: Center for Naval Analyses, November 1980.

Lockman, R.F. Success chances of recruits entering the Navy, SCREEN. (CNS 1086). Arlington, Virginia: Center for Naval Analyses, February 1977.

Matthews, W.T. Quality of Marines: Test scores, personal data, and performance. (First Term Enlisted Attrition - Vol. I: Papers). Washington, D.C.: Smithsonian Institute, June 1977.

Mehrens, W.A. & Lehmann, J.J. Measurement and evaluation in educational psychology. New York: Holt, Reinhart & Winston, 1978.

Resnick, L.B. & Glaser, R. "Problem solving and intelligence." In L.B. Resnick (Ed.), The nature of intelligence, Hillsdale, New Jersey: Erlbaum, 1976.

Sellman, W.S. The effect of mental set on job checklist information. Unpublished Doctoral Dissertation. Purdue University, 1968.

Sinaiko, H.W. & Scheflen, K.C. Attrition in the armed services of Canada, the UK, and the U.S.: A collaborative study. (TTCP(U) Technical Panel UTP-3, Military Manpower Trends). The Technical Cooperation Program, October 1980.

Thorndike, R.L. Personnel selection. New York: John Wiley & Sons, 1949.

Thorndike, R.L. & Hagen, E.P. Measurement and evaluation in psychology and education (4th Ed.). John Wiley & Sons, 1977.

U.S. Department of the Army. Guidelines for development of Skill Qualification Tests. (TRADOC Pamphlet 351-2). Fort Monroe, Virginia: Headquarters, United States Army Training and Doctrine Command, April 1980.

U.S. Office of Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics). Aptitude testing of recruits. (A report to the House Committee on Armed Services). Washington, D.C.: July 1980 (a).

U.S. Office of Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics). Implementation of new armed services Vocational Aptitude Battery and actions to improve the enlistment standards process. (A report to the House and Senate Committees on Armed Services). Washington, D.C.: December 31, 1980 (b).

Valentine, L.D. Prediction of Air Force technical training success from ASVAB and educational background. (AFHRL-TR-77-18). Lackland AFB, Texas: Air Force Human Resources Laboratory, May 1977.

Vitola, B.M., Guinn, N. & Wilbourn, J.M. Impact of various enlistment standards on the procurement-training system. (AFHRL-TR-77-16). Lackland Air Force Base, Texas: Air Force Human Resources Laboratory, April 1977.

Vitola, B.M., Mullins, C.J. & Croll, P.R. Validity of Armed Services Vocational Aptitude Battery, Form 1, to predict technical school success. (AFHRL-TR-73-7). Lackland Air Force Base, Texas: Air Force Human Resources Laboratory, July 1973.

Wesman, A.G. Writing the test item. In R.L. Thorndike, (Ed.), Educational Measurement (2nd Ed.), Washington, D.C.: American Council on Education, 1971.

PRI Program Resources, Inc.
6001 Montrose Road
Suite 405
Rockville, Maryland 20852 • (301) 984-9020

Appendix A

ASVAB APTITUDE COMPONENTS

Table 20

Selected Aptitude Composites for ASVAB Forms 6 and 7

Service	Comp.	ASVAB Subtests: Word Knowl.	Arith. Reas.	Space Perc.	Shop Info.	Atten. Det.	Elect. Info.	Mech. Comp.	Gener. Sci.	Math. Know.	Auto Info.	Numer. Oper.	Army Class. Inven.
	AFQT	X	X	X									
Army & MC	CO		X	X	X	X							X
Army	SC	X	X	X				X					
Army	EL		X		X		X	X					X
Marine Corps	EL		X				X		X	X			
Army & MC	GT	X	X										
Army	MM				X		X			X	X		X
Army	CL	X	X			X						X	

Table 21

Selected Aptitude Composites for ASVAB Forms 8, 9, and 10

ASVAB Subtests

Service	Comp.	Verbal: Word Knowl.	Verbal: Para. Comp.	Arith. Reas.	Numer. Oper.	Coding Speed	Mech. Comp.	Auto/Shop Info.	Math. Knowl.	Electro. Info.	Gen. Sci.
	AFQT	X	X	X	X						
MC	CO	X	X		X			X			
Army	CO			X		X	X	X			
Army	SC	X	X		X	X		X			
Army & MC	EL			X					X	X	X
Army & MC	GT	X	X	X							
Army	MM				X		X	X		X	
Army	CL	X	X		X	X					

Appendix B

DATA COLLECTION VISITS

Table 22. Schedule of Data Collection Visits

Table 22

Schedule of Data Collection Visits

MOS	JOB	Proponent School	Location	Week of Visit	Service
0311	Infantryman	Camp Pendleton	San Clemente, CA	1/19/81	Marine Corps
11B	Infantryman	Ft. Benning	Columbus, GA	3/23/81	Army
11C	Indirect Fire Infantryman	Ft. Benning	Columbus, GA	3/23/81	Army
19E	Armor Crewman	Ft. Knox	Louisville, KY	4/20/81	Army
05C	Radio Teletypewriter Operator	Ft. Gordon	Augusta, GA	3/ 9/81	Army
31M	Multichannel Communications Operator	Ft. Gordon	Augusta, GA	3/ 9/81	Army
2841	Ground Radio Repair	Camp Twentynine Palms	Twentynine Palms, CA	1/26/81	Marine Corps
67N	Utility Helicopter Repairer	Ft. Rucker	Dothan, AL	5/ 4/81	Army
73C	Finance Specialist	Ft. Benjamin Harrison	Indianapolis, IN	4/ 6/81	Army
75B	Personnel Administration Specialist	Ft. Benjamin Harrison	Indianapolis, IN	4/ 6/81	Army

Appendix C

SQT COMPONENT MIX

Table 23. Recommended Component Mix for SQT Designed to Test a Skill Level 1 Soldier

Table 23

Recommended Component Mix for SQT
Designed to Test a Skill Level 1 Soldier

	Number of Tasks in Each Component			
Type MOS	SC	HOC	JSC	TOTAL
Combat Arms	0-6	13-17	11	24-34
Combat Support MOS	4-9	10-15	11	25-35
Combat Service Support	7-12	7-12	11	25-35

Appendix D

RELATIONSHIP OF AFQT AND APTITUDE COMPOSITE SCORES TO SQT PERFORMANCE

- Figures 3-10 and 19-26 in Appendix D show the relationship of AFQT and aptitude composite scores to Skill Qualification Test performance for the eight Army MOSs chosen for this study.
- Figures 11 through 18 and 27 through 34 show the breakdown by education.
- Figures 35 through 50 look at the relationship between AFQT/aptitude composite and SQT score as a function of different racial/ethnic groups.

Figure 3
SQT Performance as a Function of AFQT Category for
MOS 11B (Infantryman)

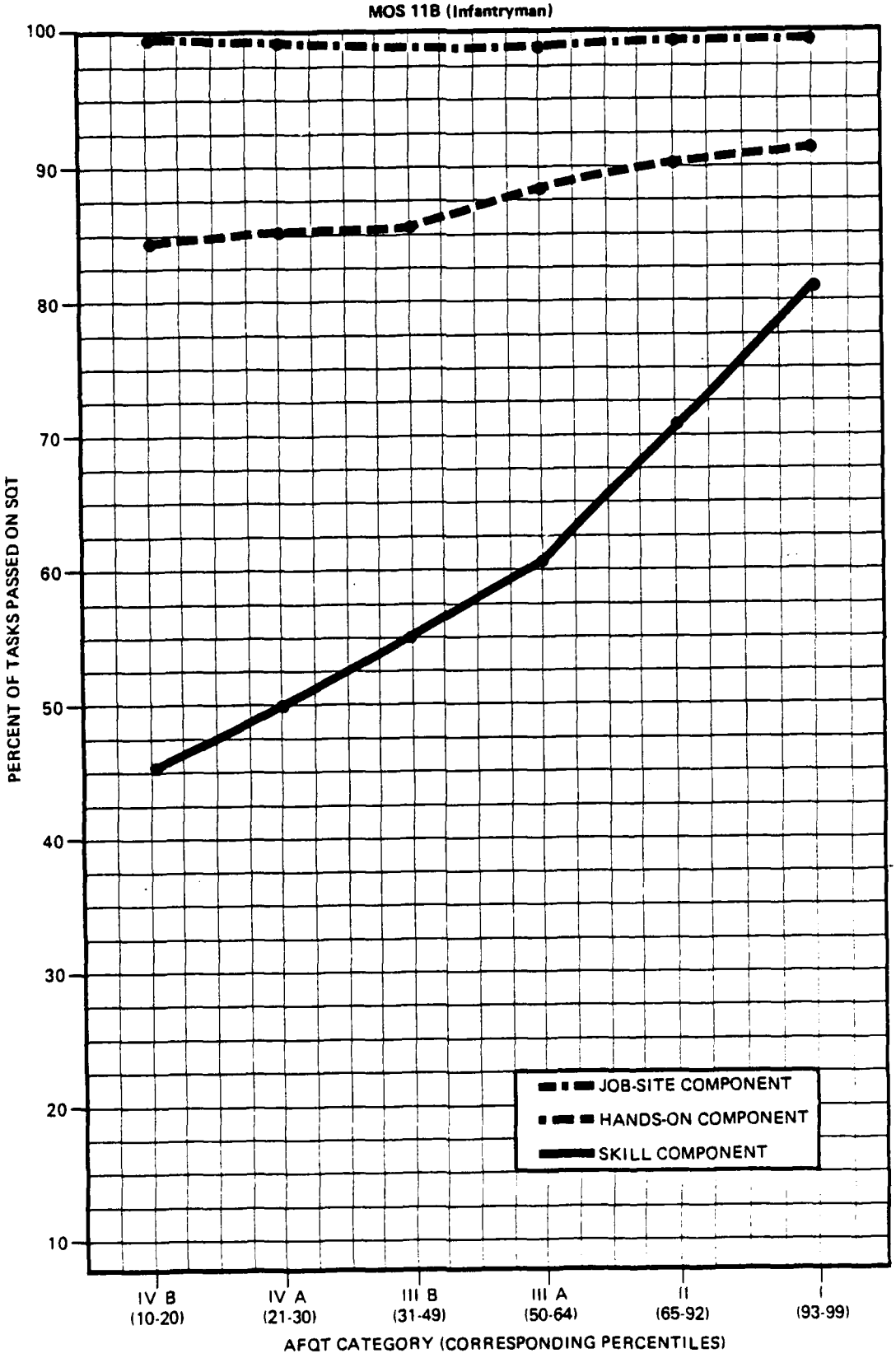


Figure 4
SQT Performance as a Function of AFQT Category for
MOS 11C (Indirect Fire Infantryman)

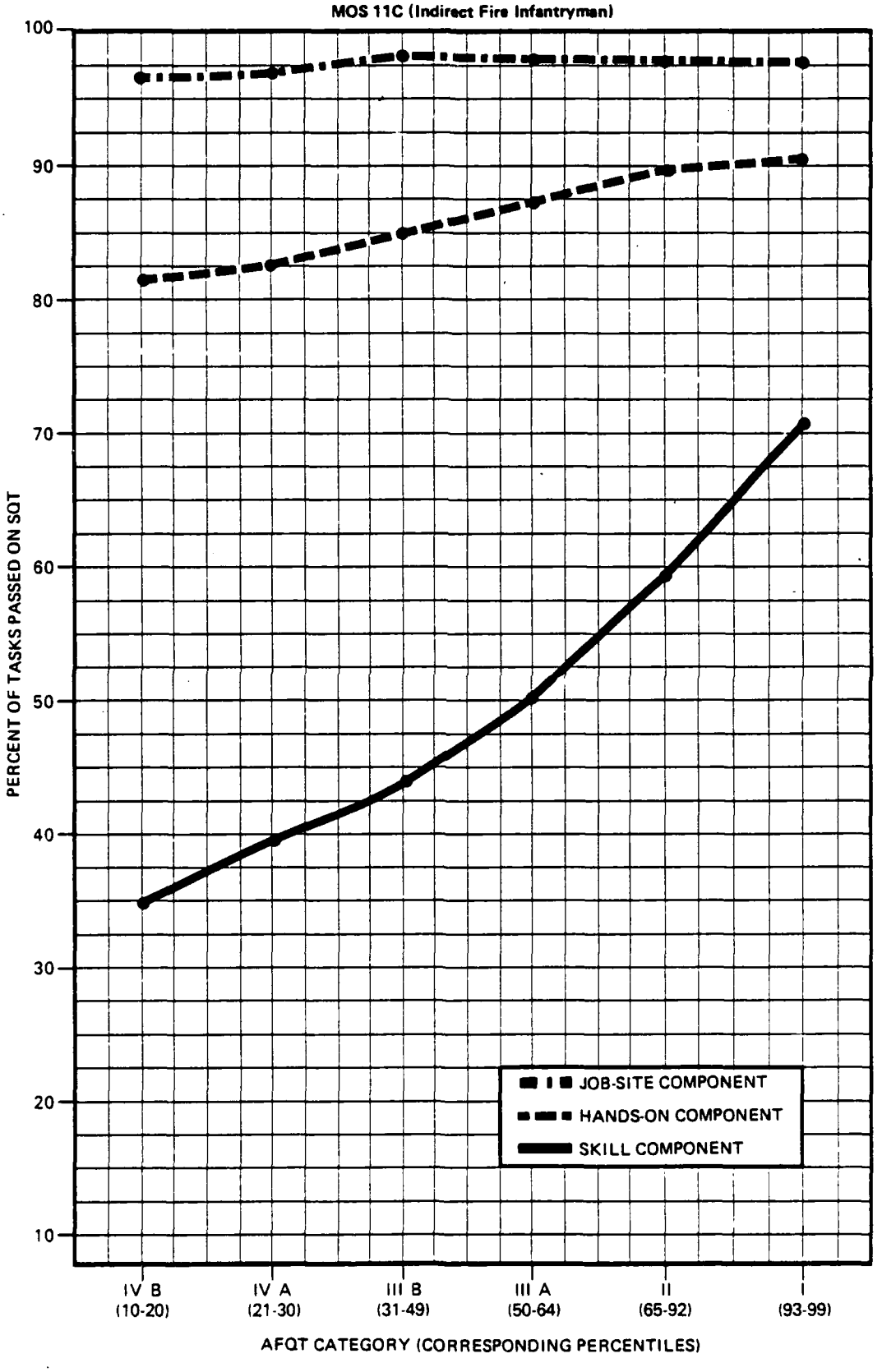


Figure 5
SQT Performance as a Function of AFQT Category for
MOS 19E (Armor Crewman)

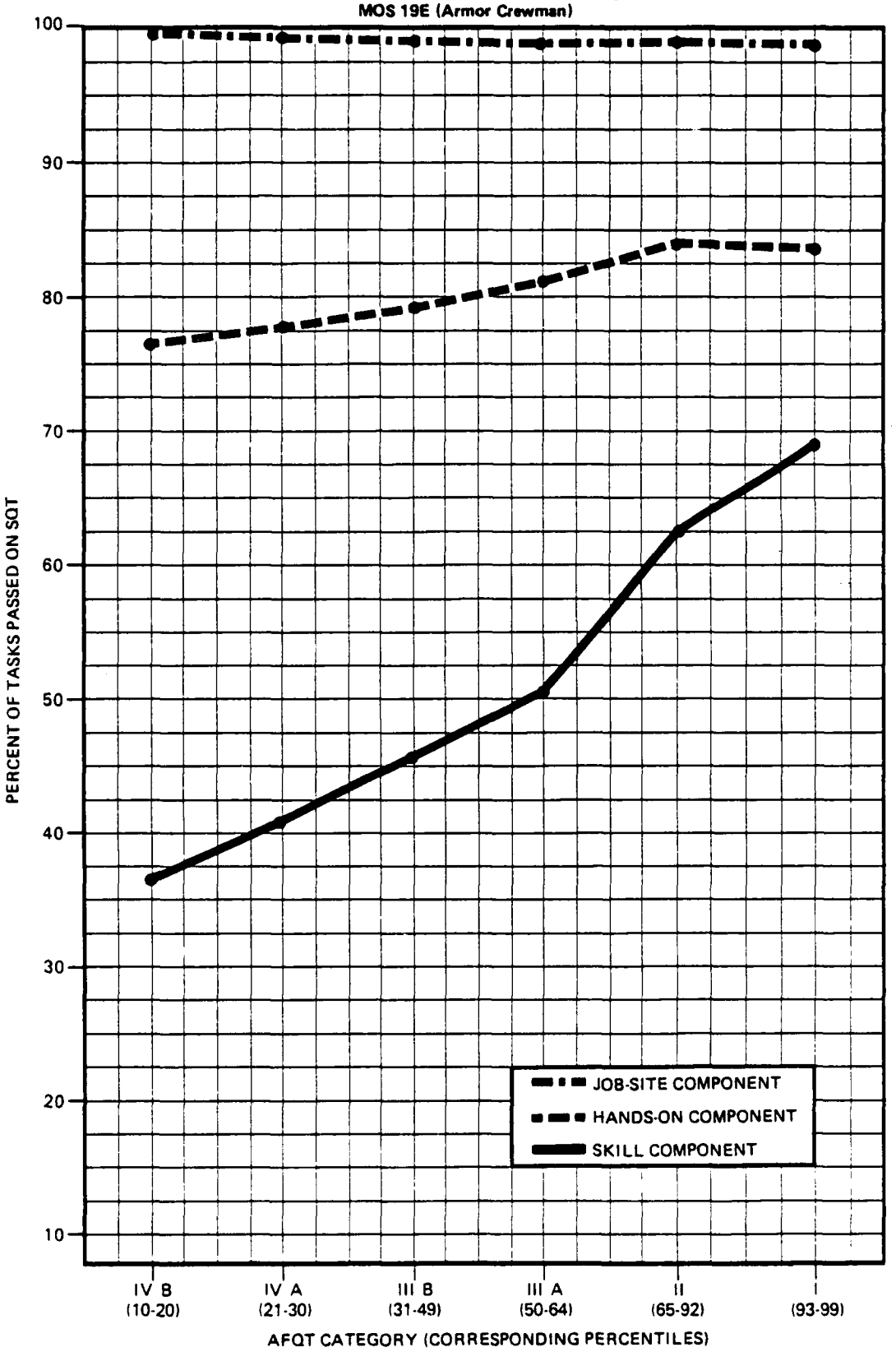


Figure 6
SQT Performance as a Function of AFQT Category for
MOS 05C (Radio Teletypewriter Operator)

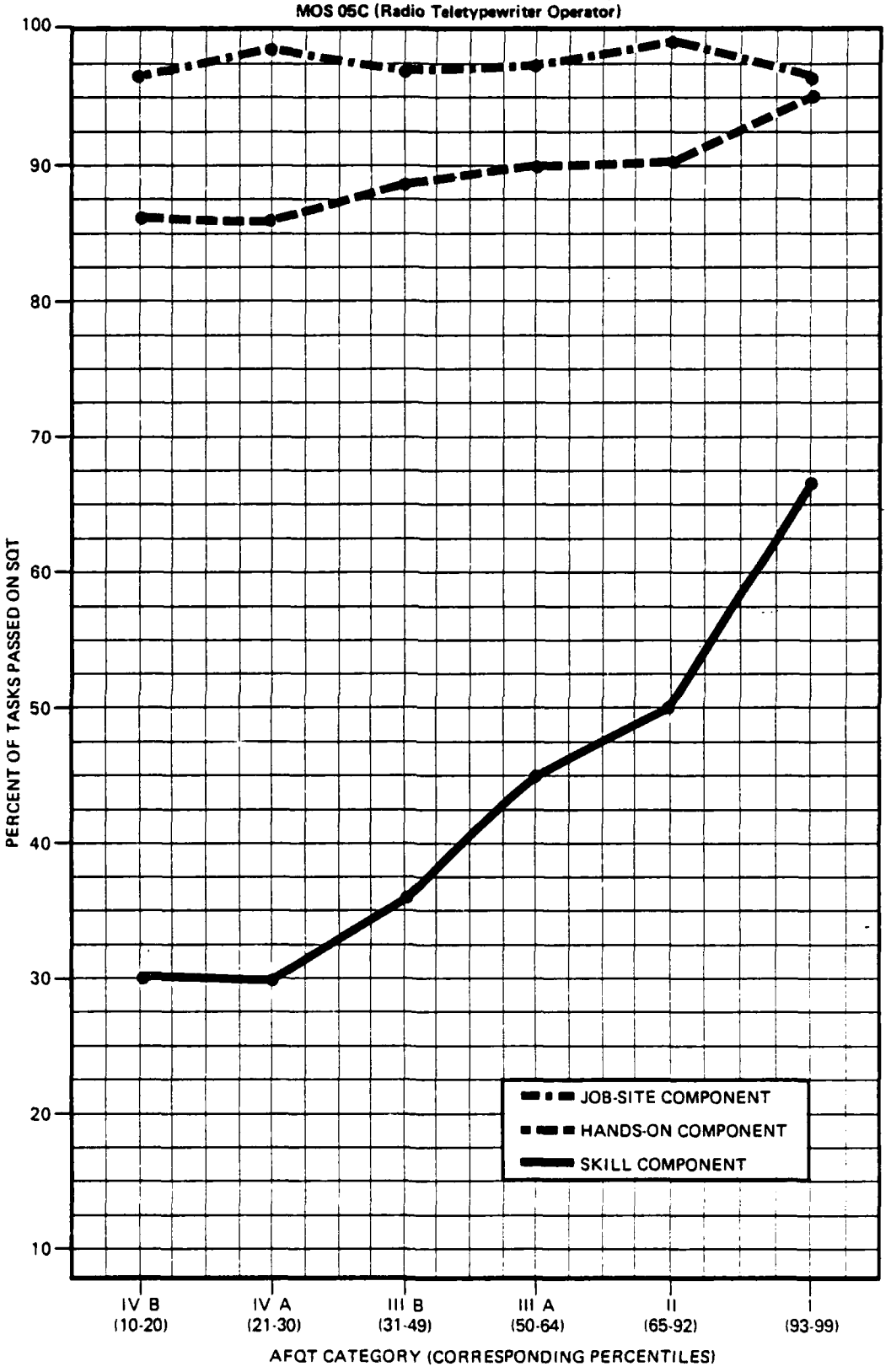


Figure 7
SQT Performance as a Function of AFQT Category for
MOS 31M (Multichannel Communications Operator)

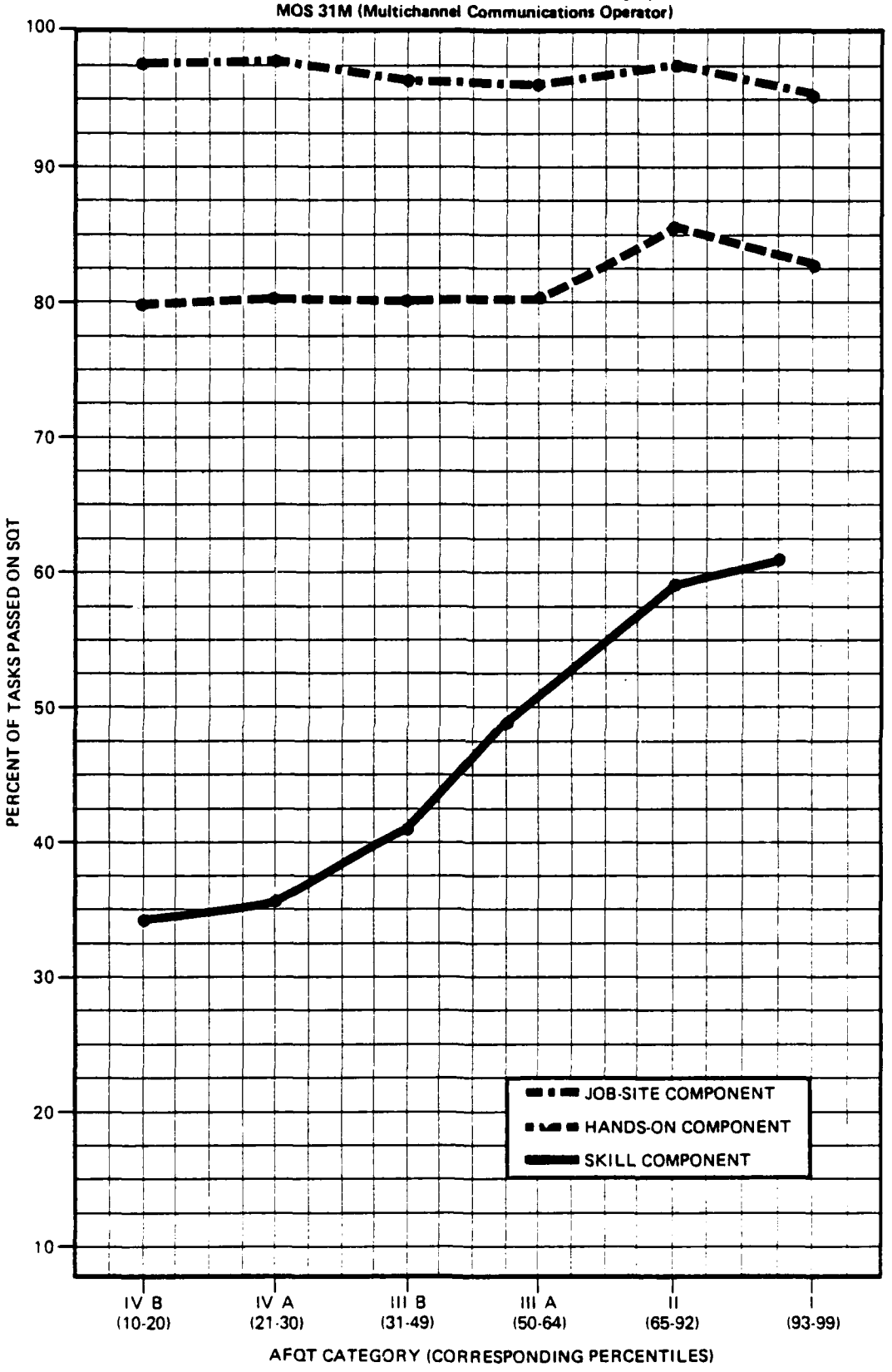


Figure 8
SQT Performance as a Function of AFQT Category for
MOS 67N (Utility Helicopter Repairer)

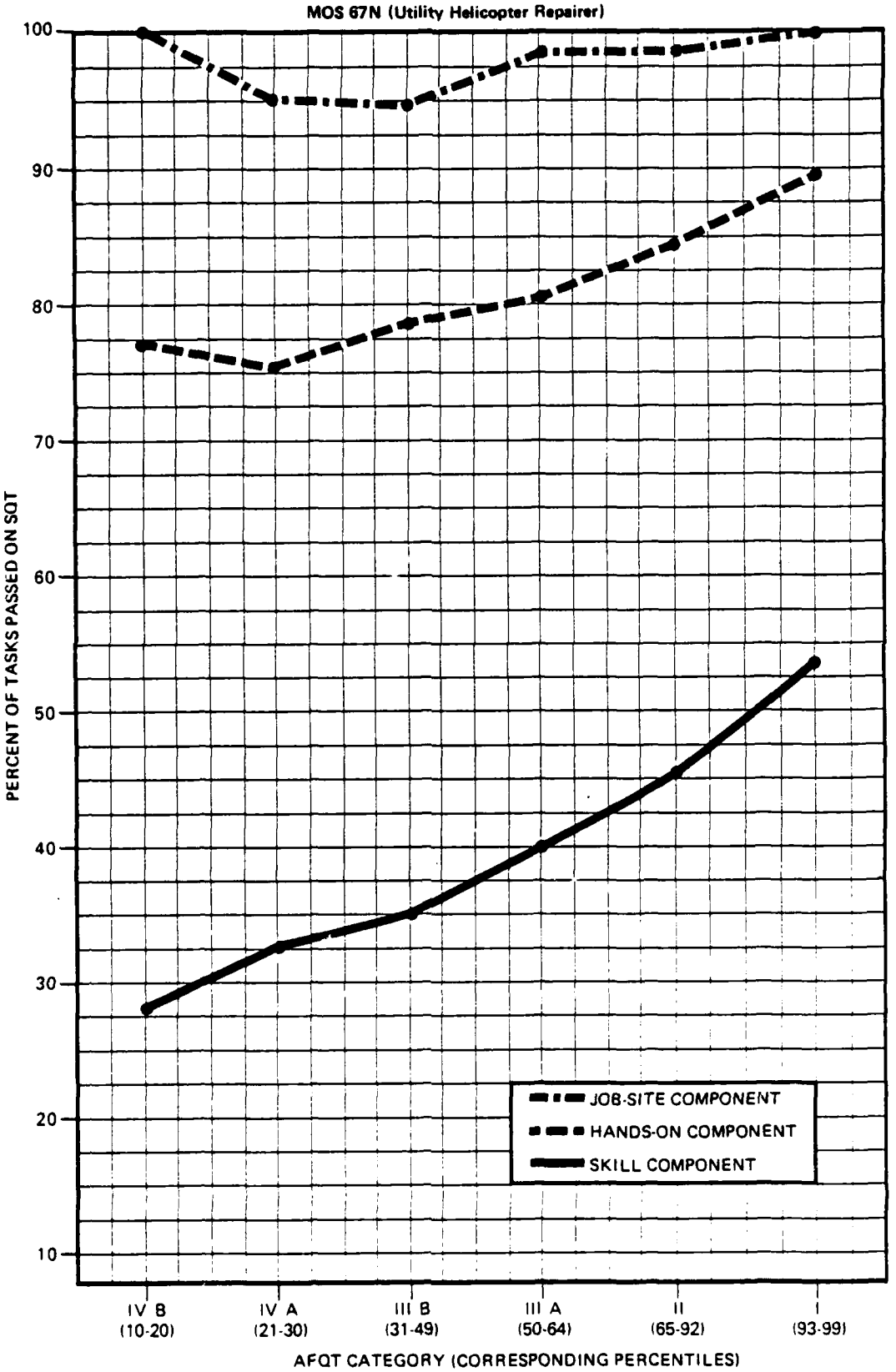
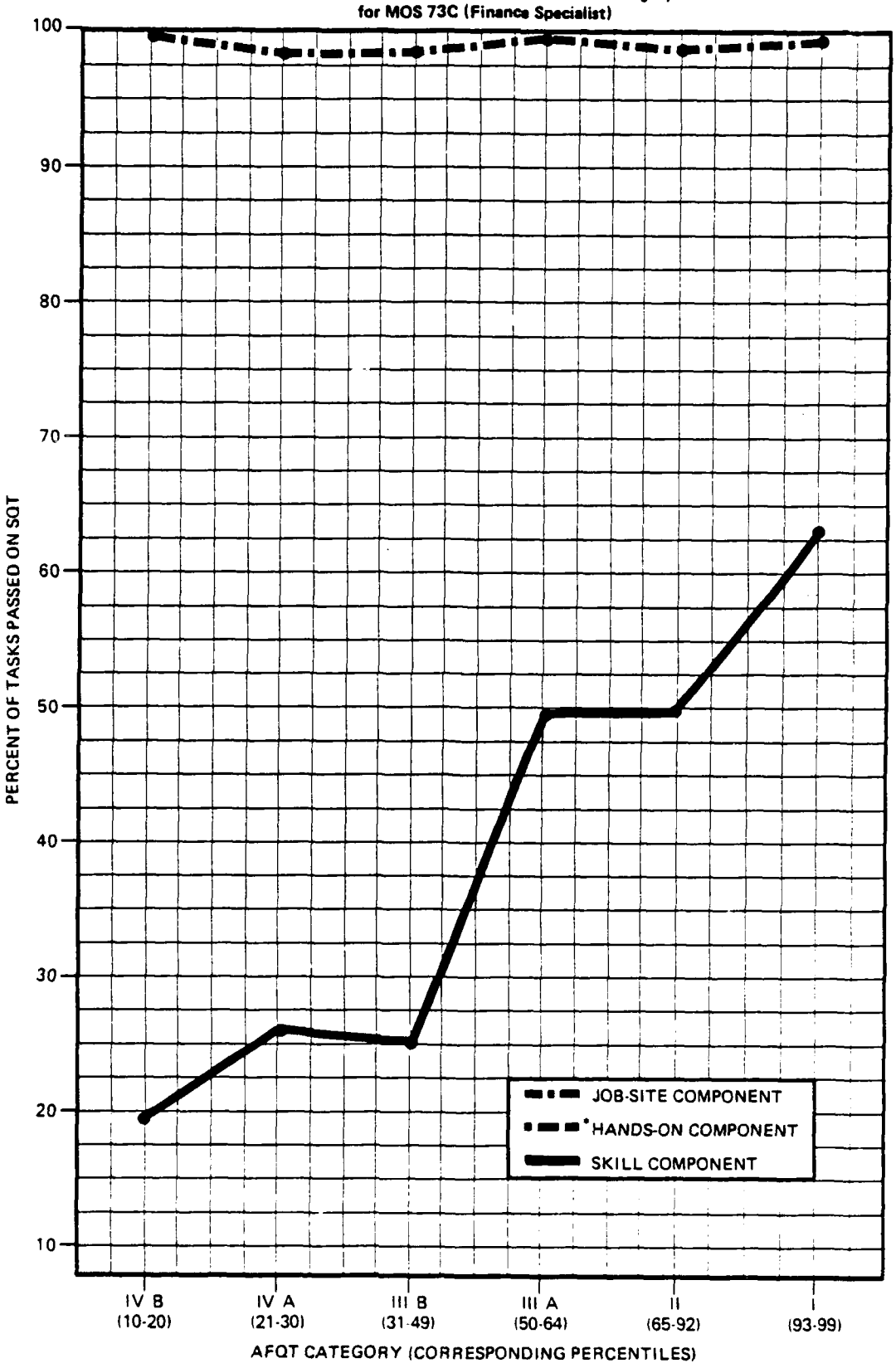


Figure 9
SQT Performance as a Function of AFQT Category
for MOS 73C (Finance Specialist)



*There was no HOC test for 73C in 1980, when most of the data was obtained.

Figure 10
SQT Performance as a Function of AFQT Category
for MOS 75B (Personnel Administration Specialist)

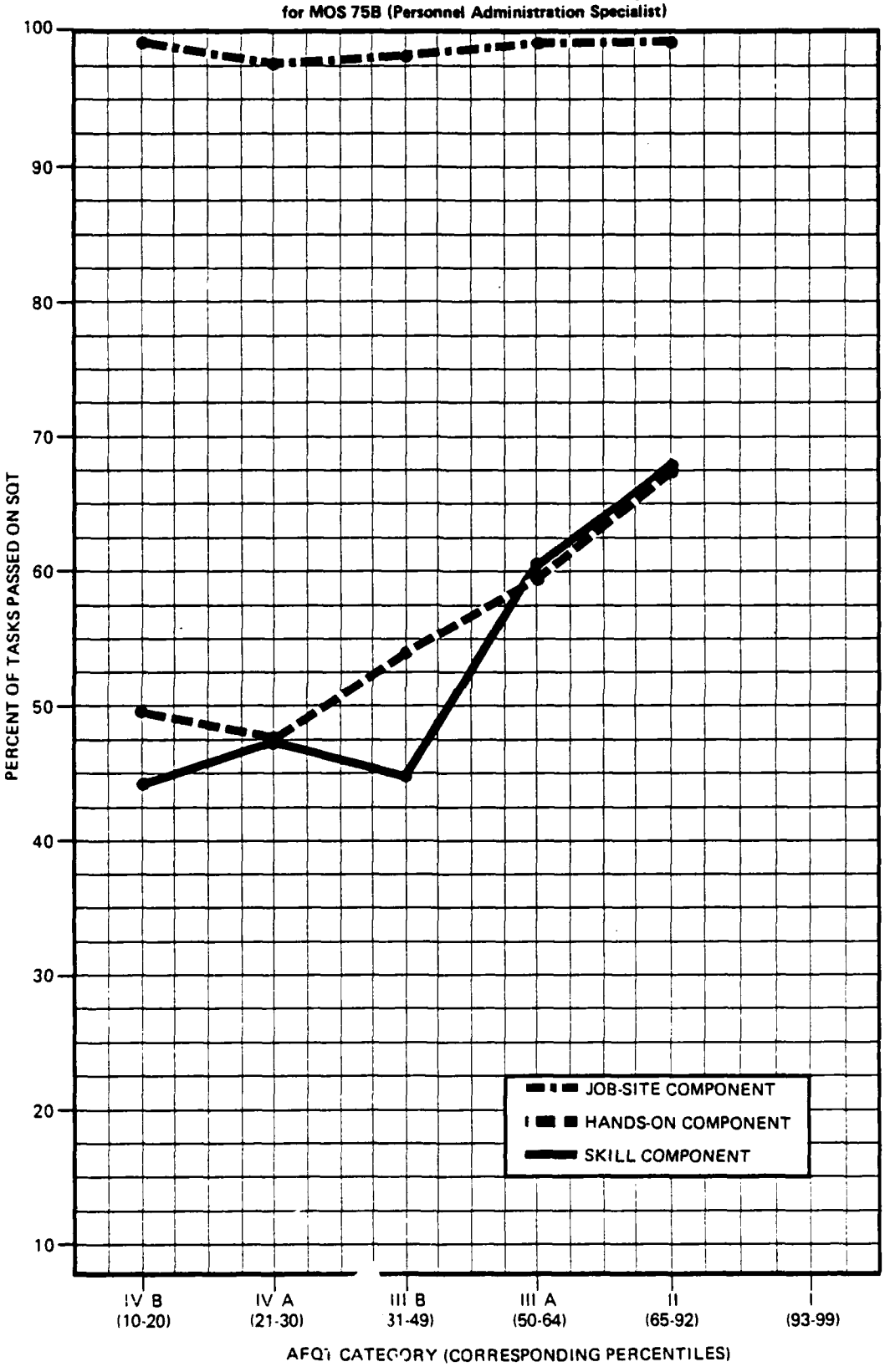


Figure 11
SQT Performance as a Function of AFQT Category and Education
for MOS 11B (Infantryman)

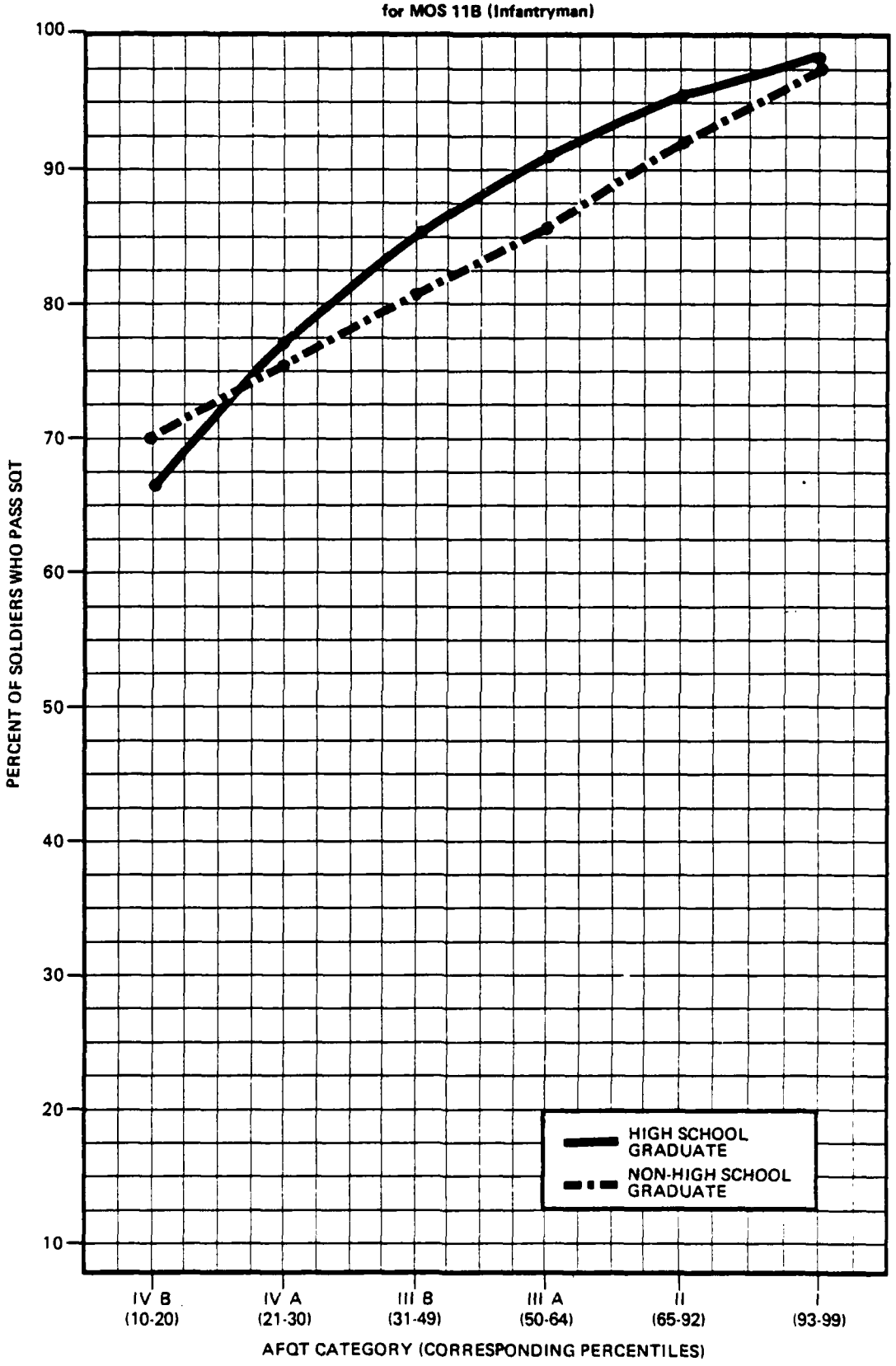


Figure 12
SQT Performance as a Function of AFQT Category and Education
for MOS 11C (Indirect Fire Infantryman)

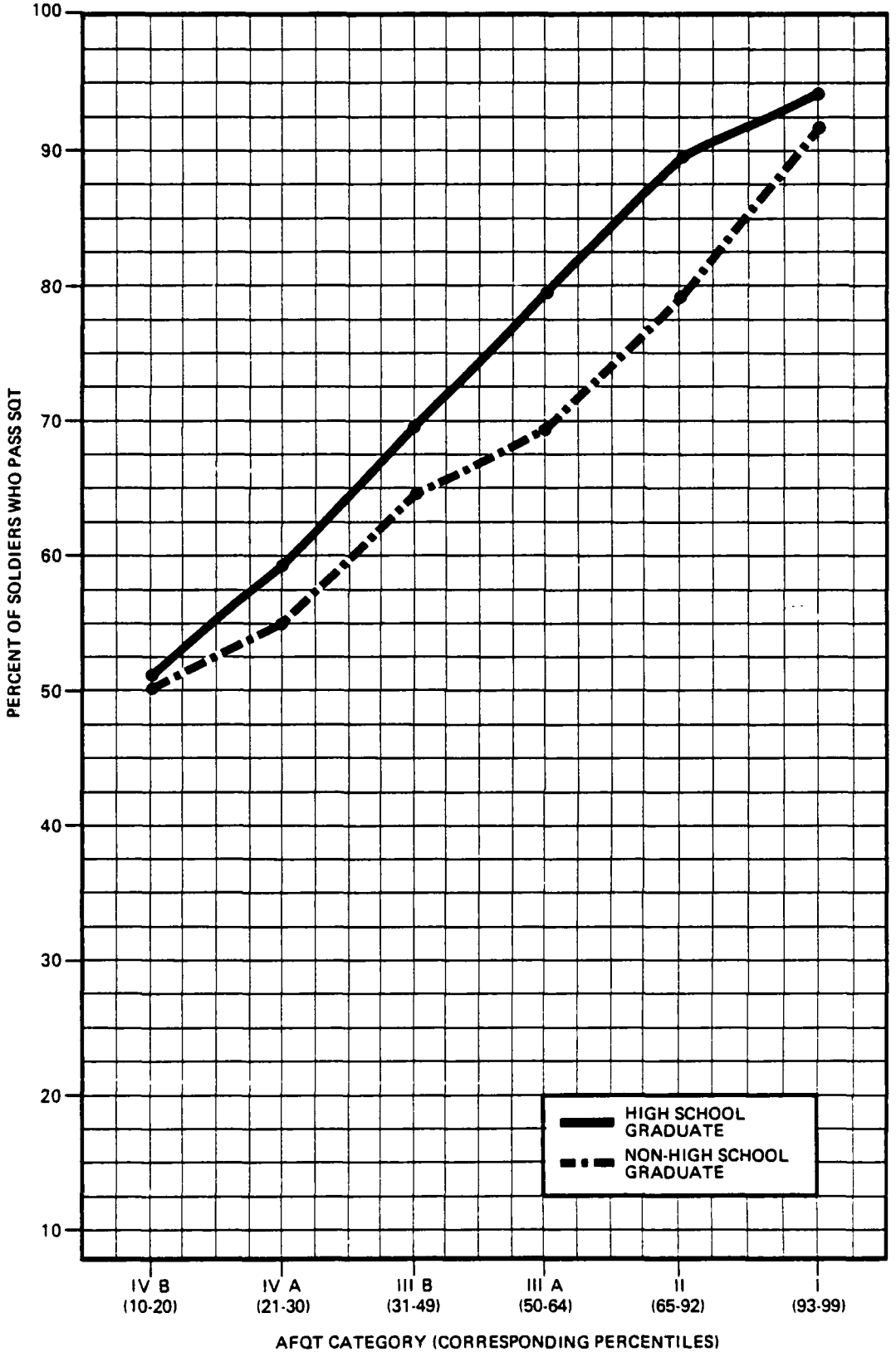


Figure 13
SQT Performnace as a Function of AFQT Category and Education
for MOS 19E (Armor Crewman)

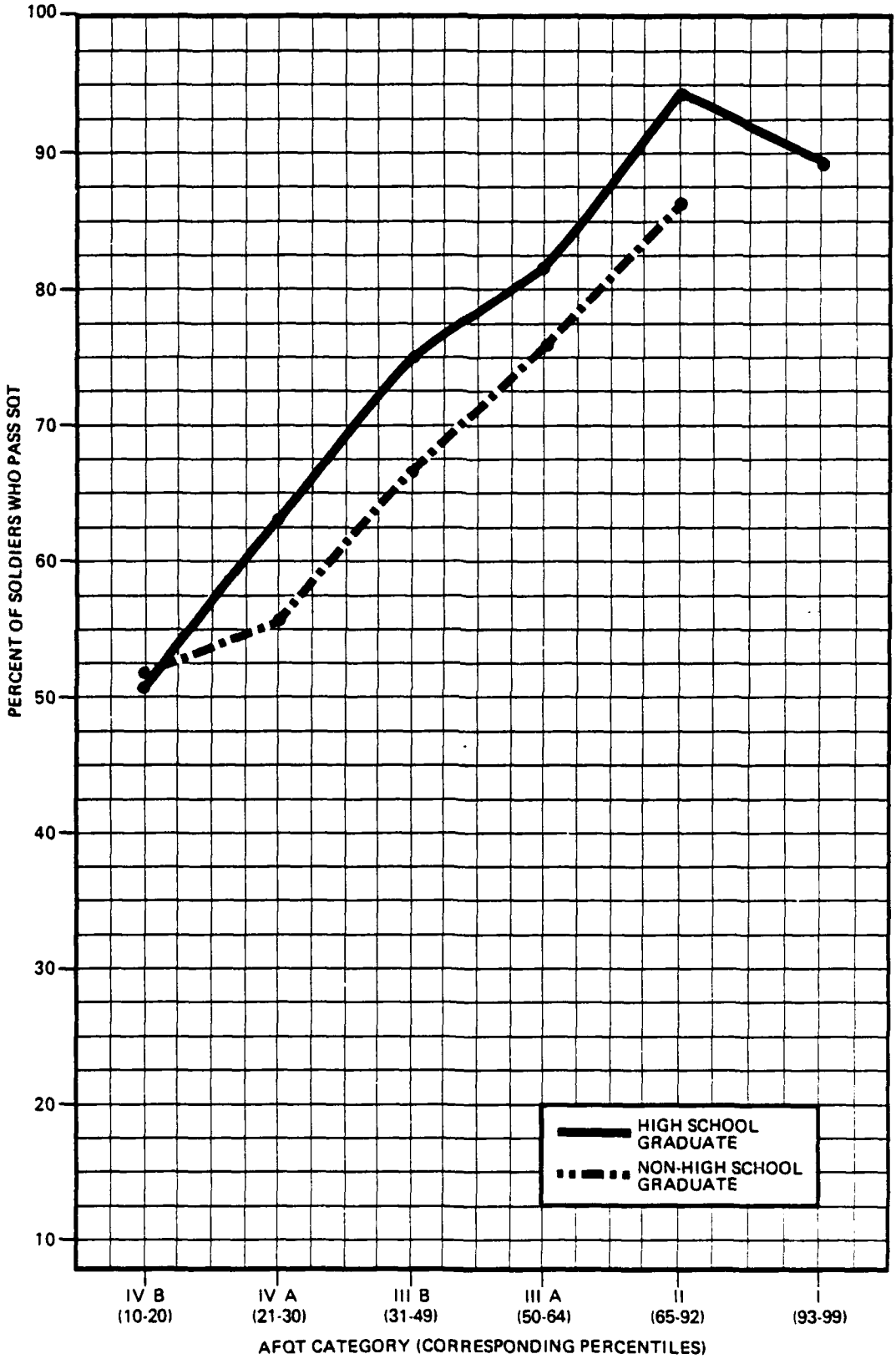


Figure 14
SQT Performance as a Function of AFQT Category and Education
for MOS 05C (Radio Teletypewriter Operator)

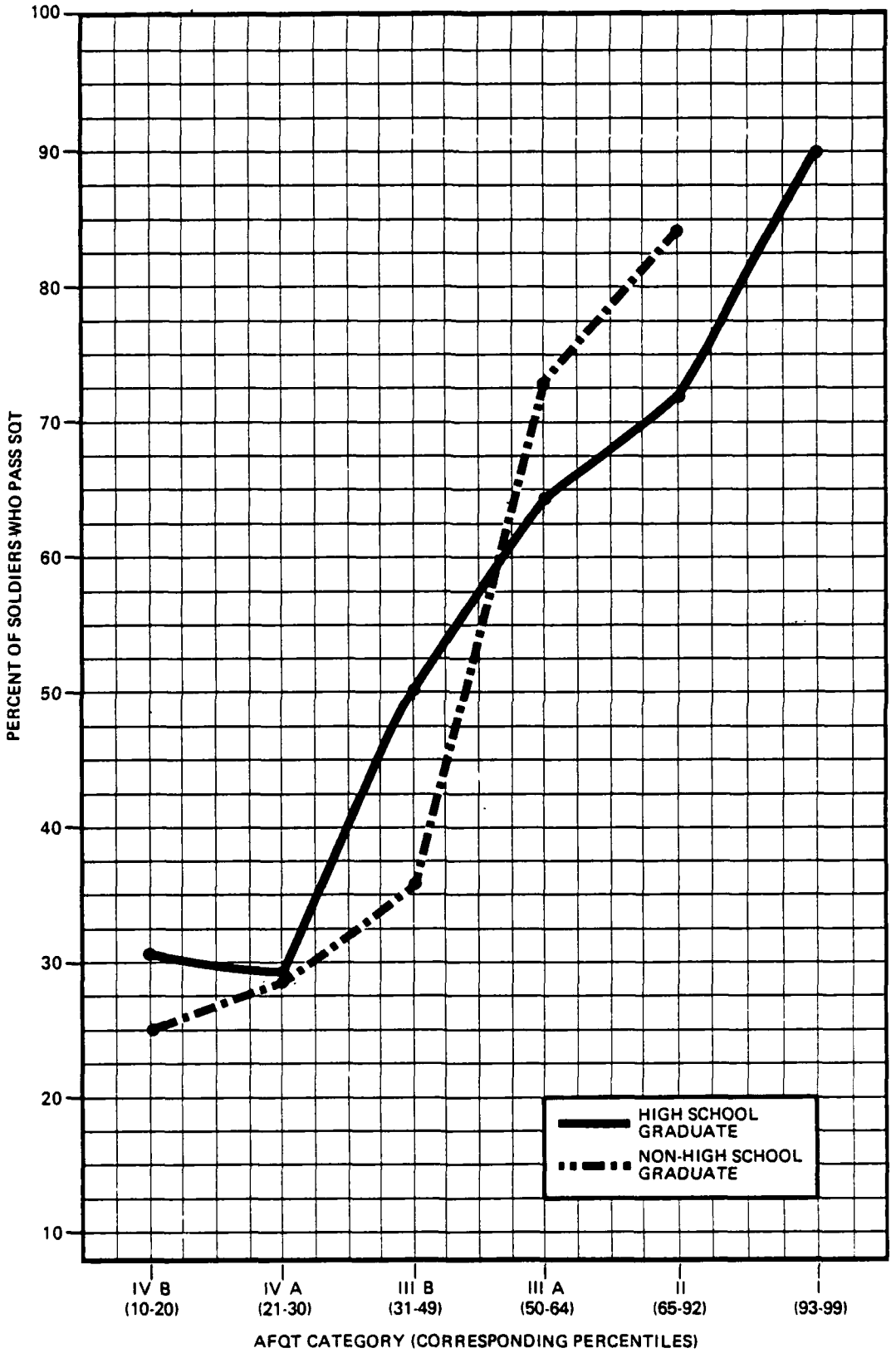


Figure 15
SQT Performance as a Function of AFQT Category and Education
for MOS 31M (Multichannel Communications Operator)

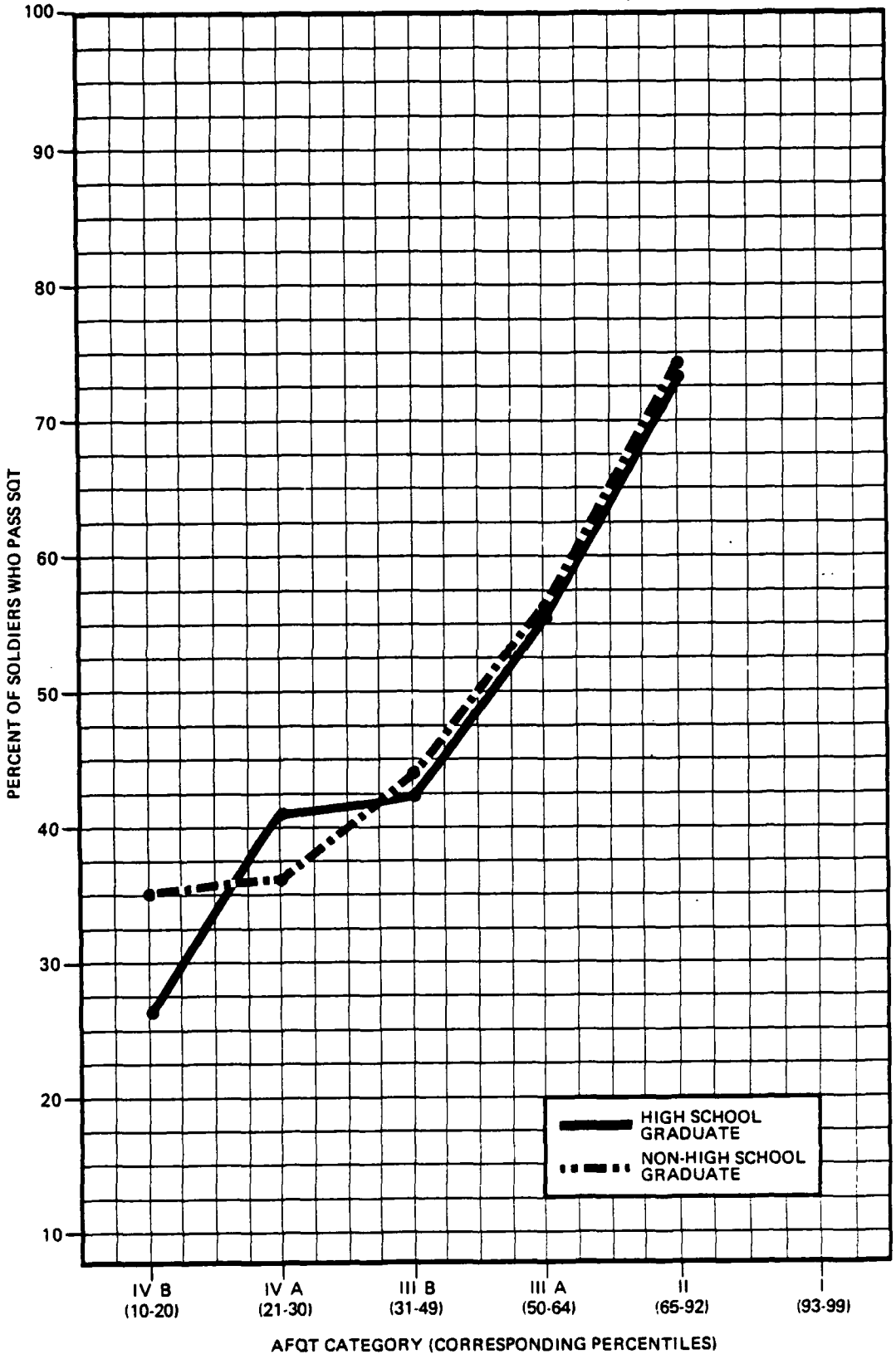


Figure 16
SQT Performance as a Function of AFQT Category and Education
for MOS 67N (Utility Helicopter Repairer)

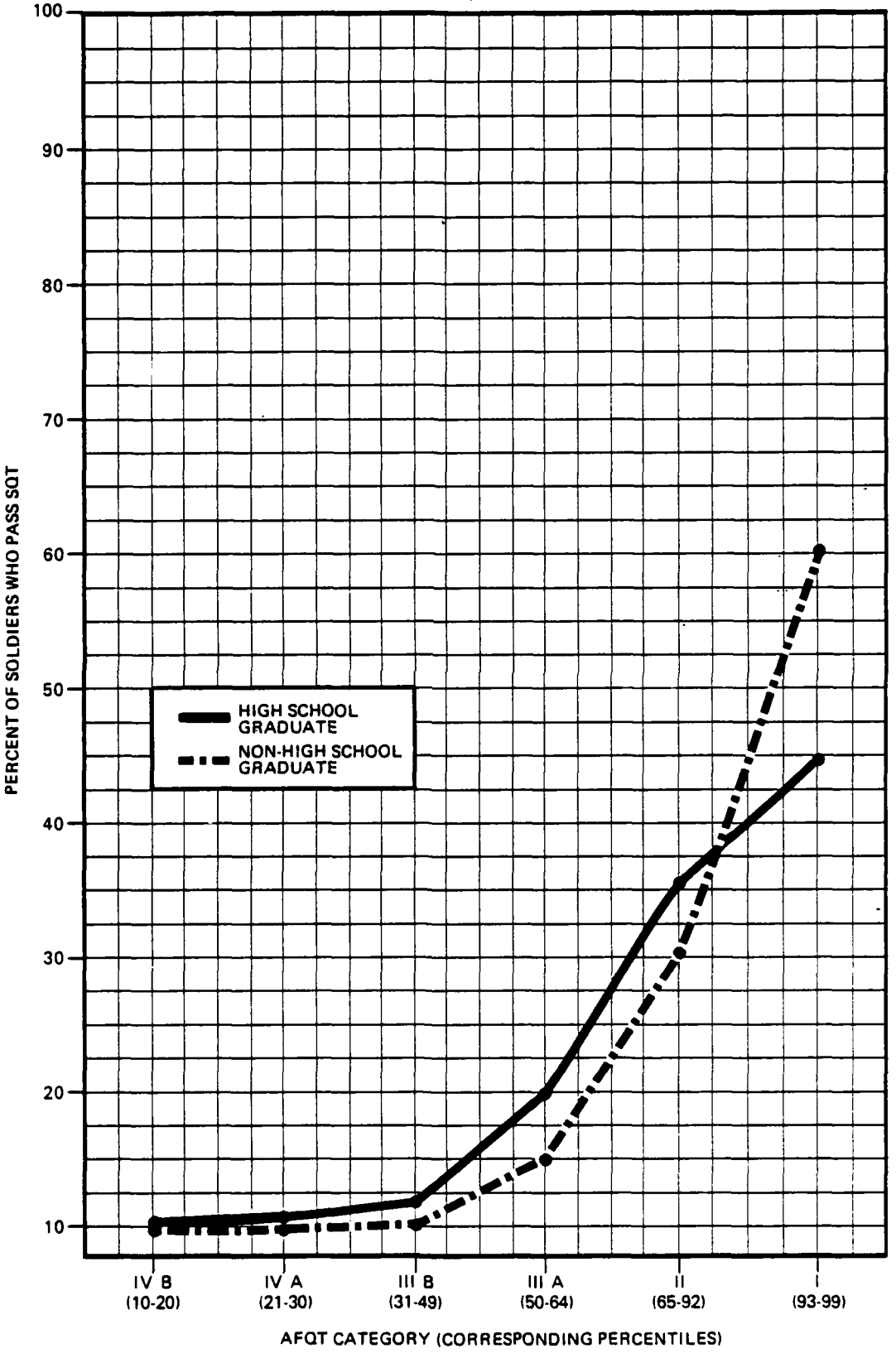


Figure 17
SQT Performance as a Function of AFQT Category and Education
for MOS 73C (Finance Specialist)

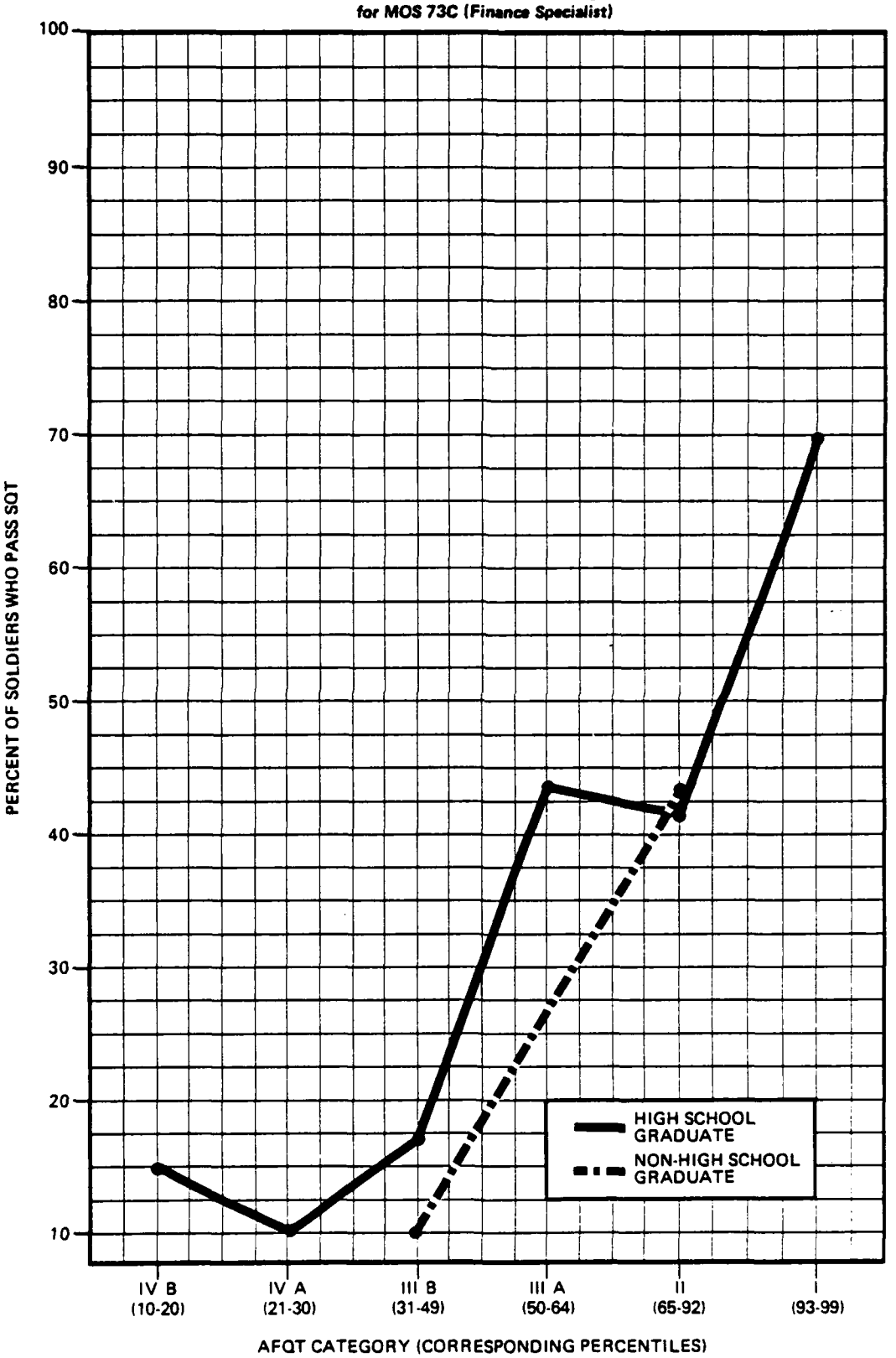


Figure 18
SQT Performance as a Function of AFQT Category and Education for MOS 75B (Personnel Administration Specialist)

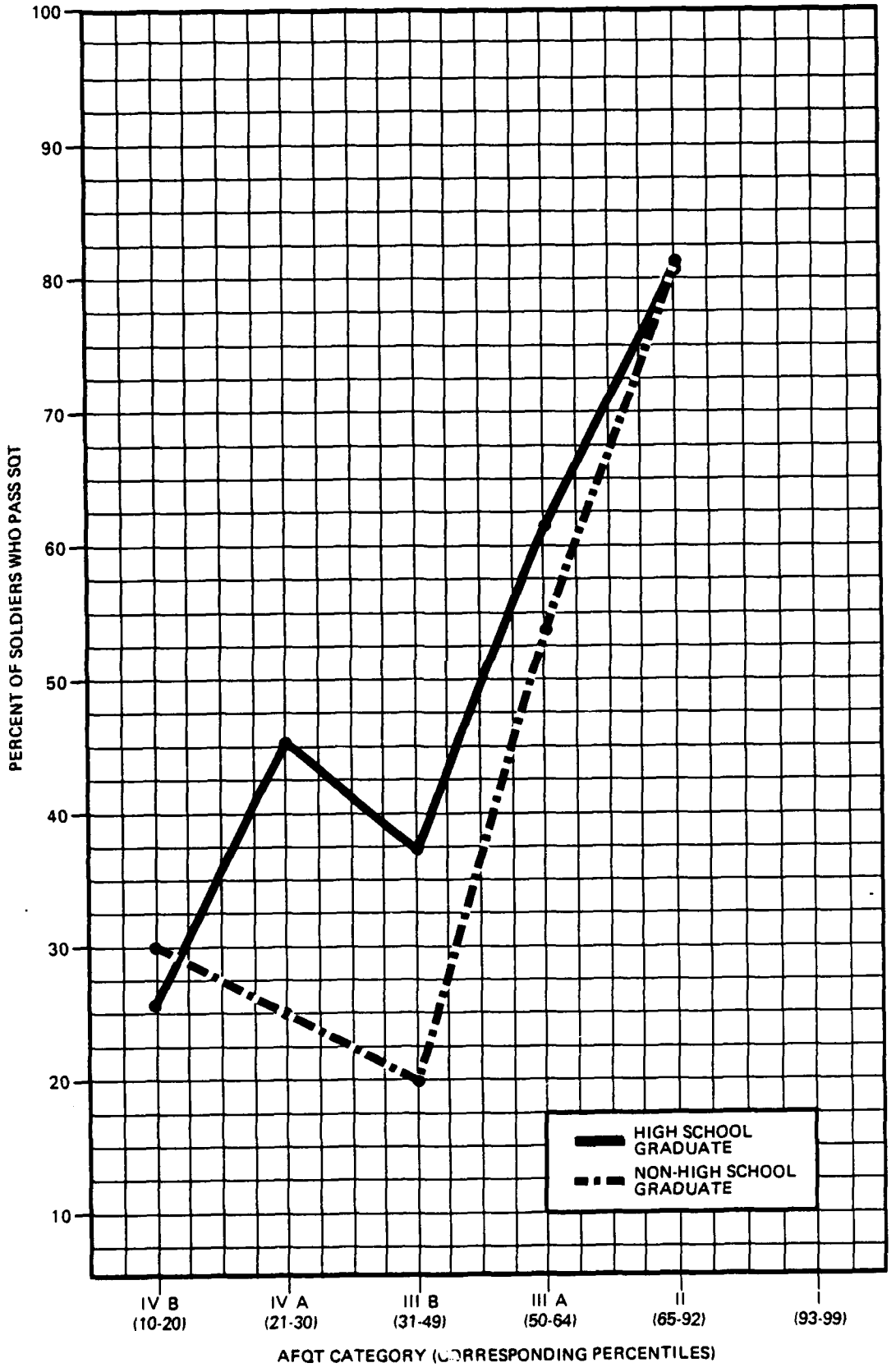


Figure 19
SQT Performance as a Function of Aptitude Composite Score
for MOS 11B (Infantryman)

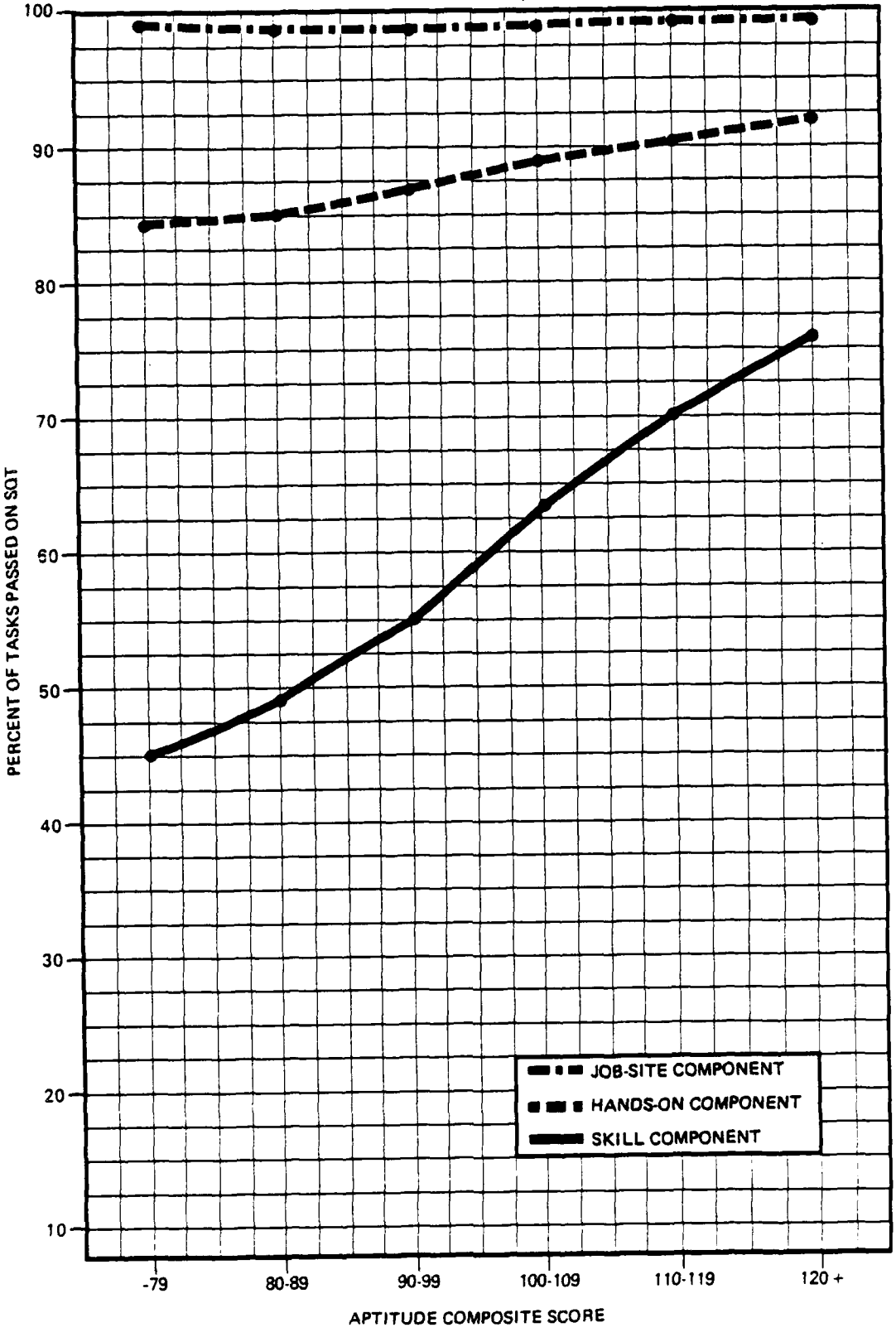


Figure 20
SQT Performance as a Function of Aptitude Composite Score
for MOS 11C (Indirect Fire Infantryman)

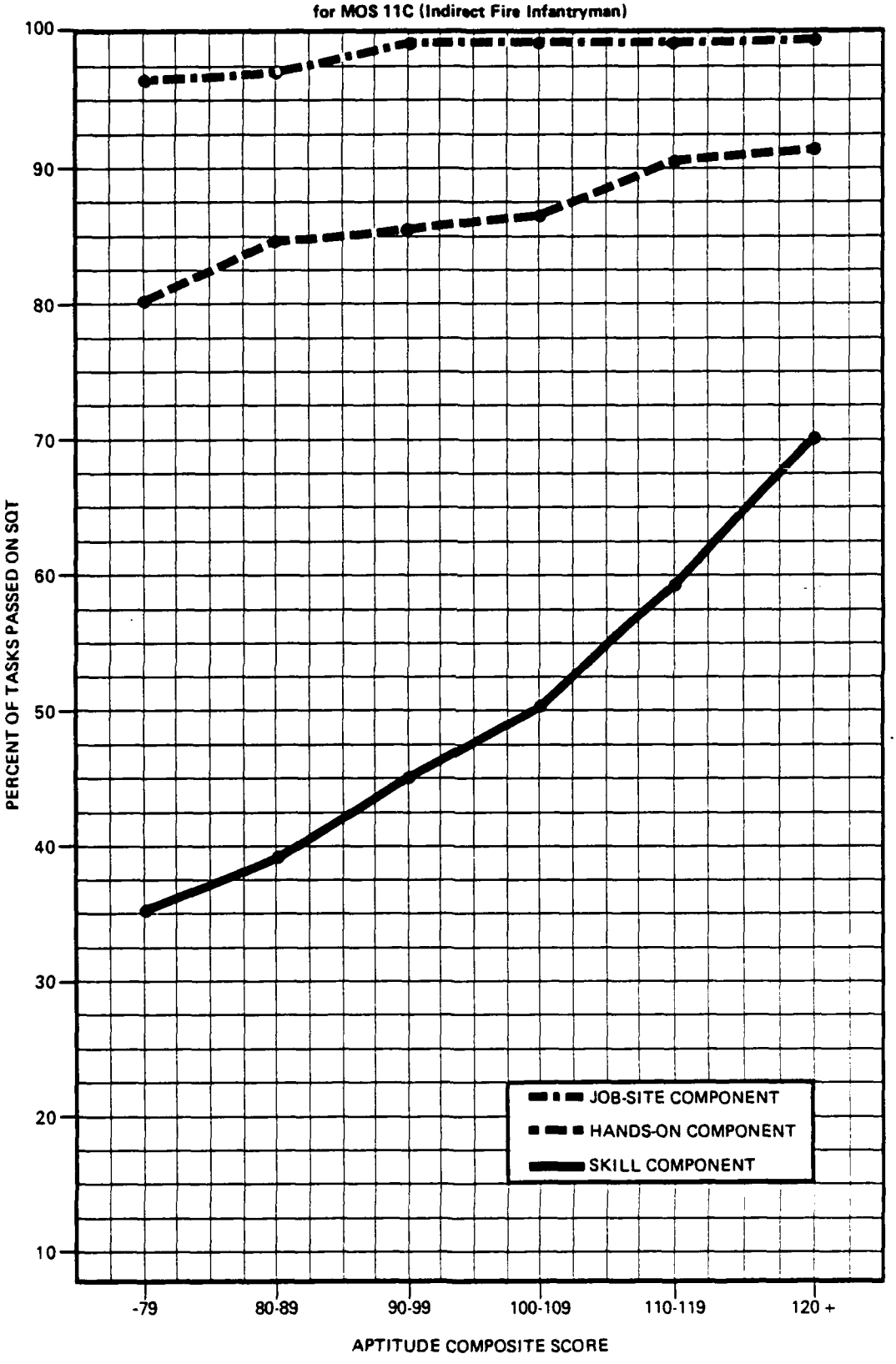


Figure 21
SQT Performance as a Function of Aptitude Composite Score
for MOS 19E (Armor Crewman)

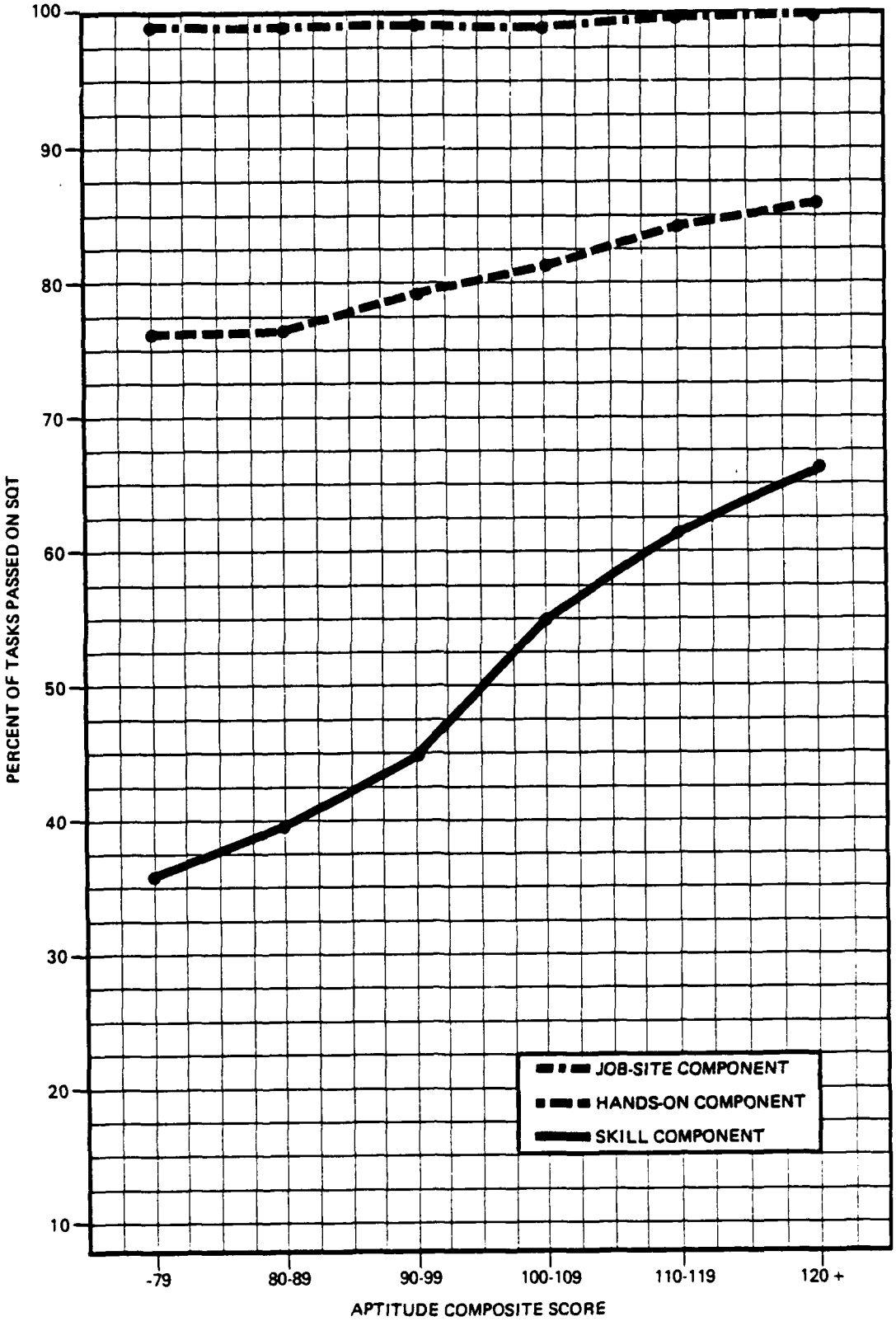


Figure 22
SQT Performance as a Function of Aptitude Composite Score
for MOS 05C (Radio Teletypewriter Operator)

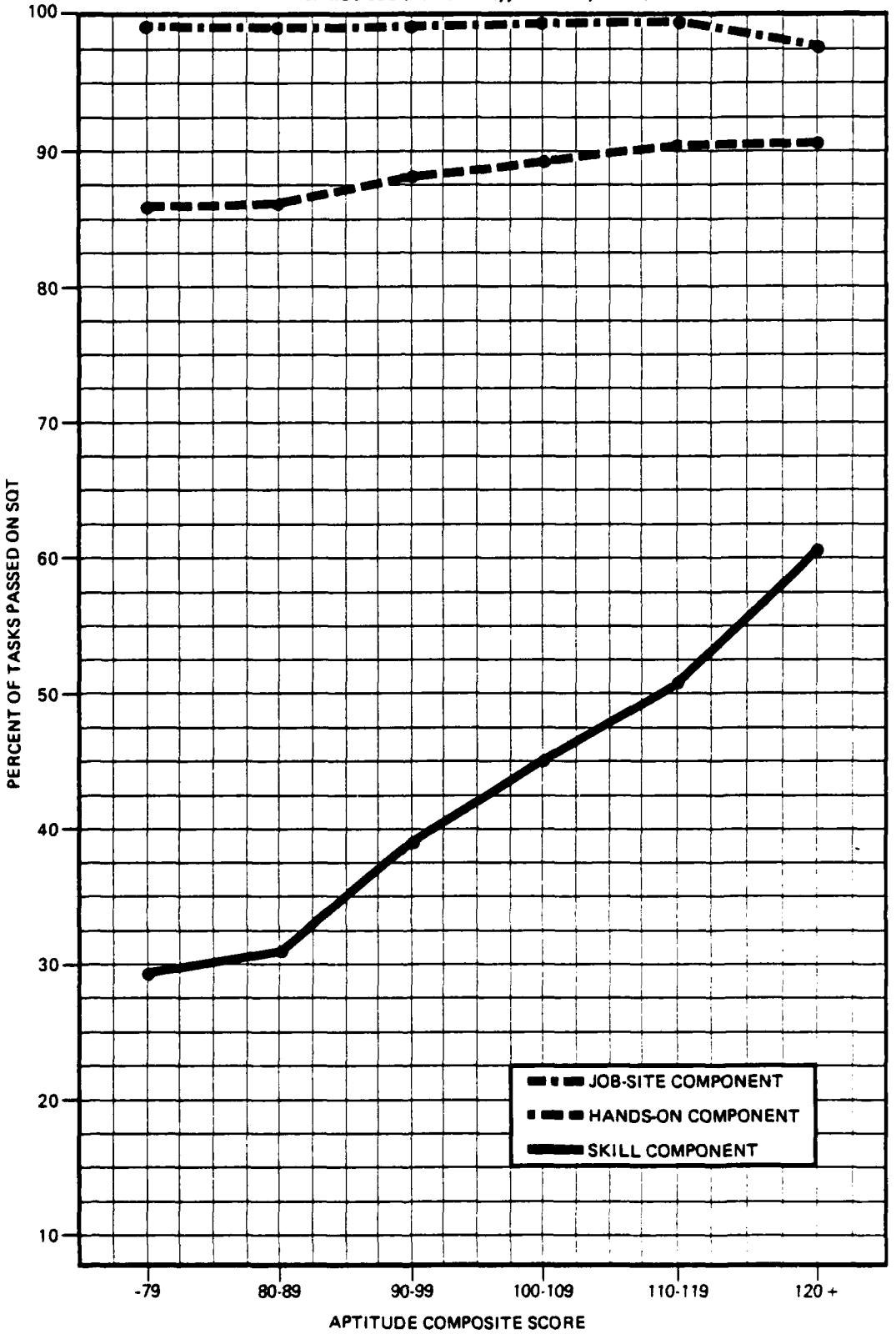


Figure 23
SQT Performance as a Function of Aptitude Composite Score
for MOS 31M (Multichannel Communications Operator)

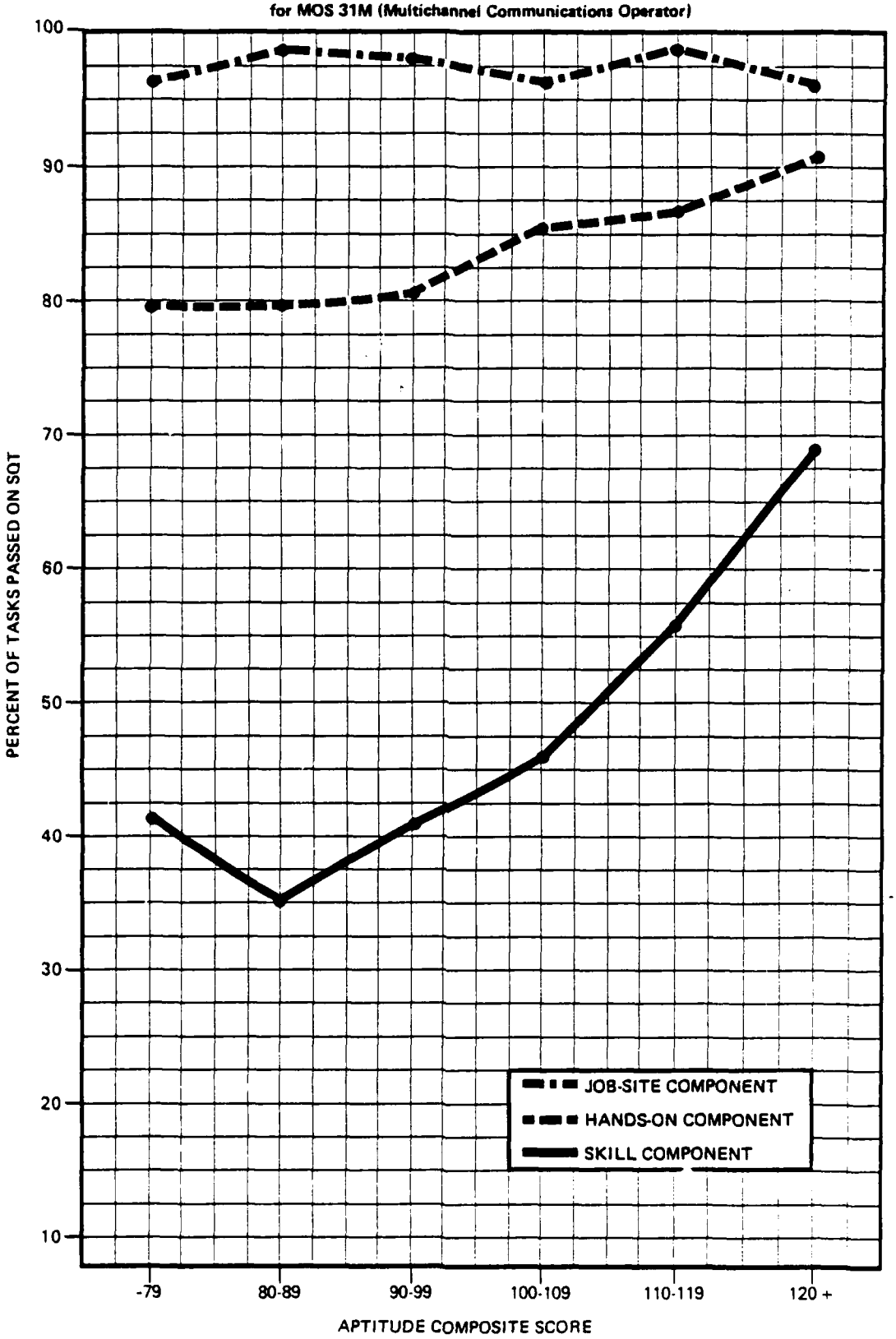


Figure 24
SQT Performance as a Function of Aptitude Composite Score
for MOS 67N (Utility Helicopter Repairer)

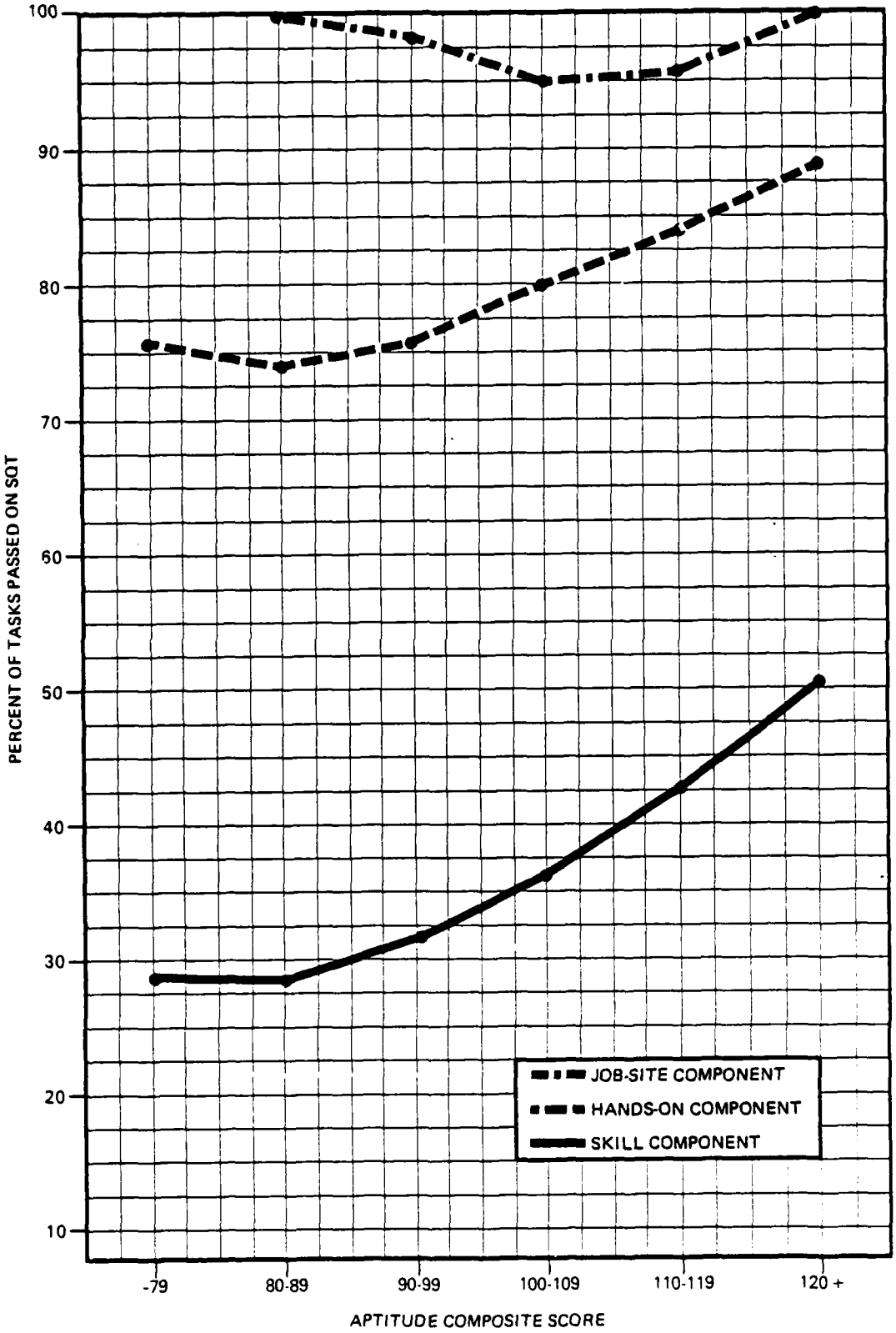
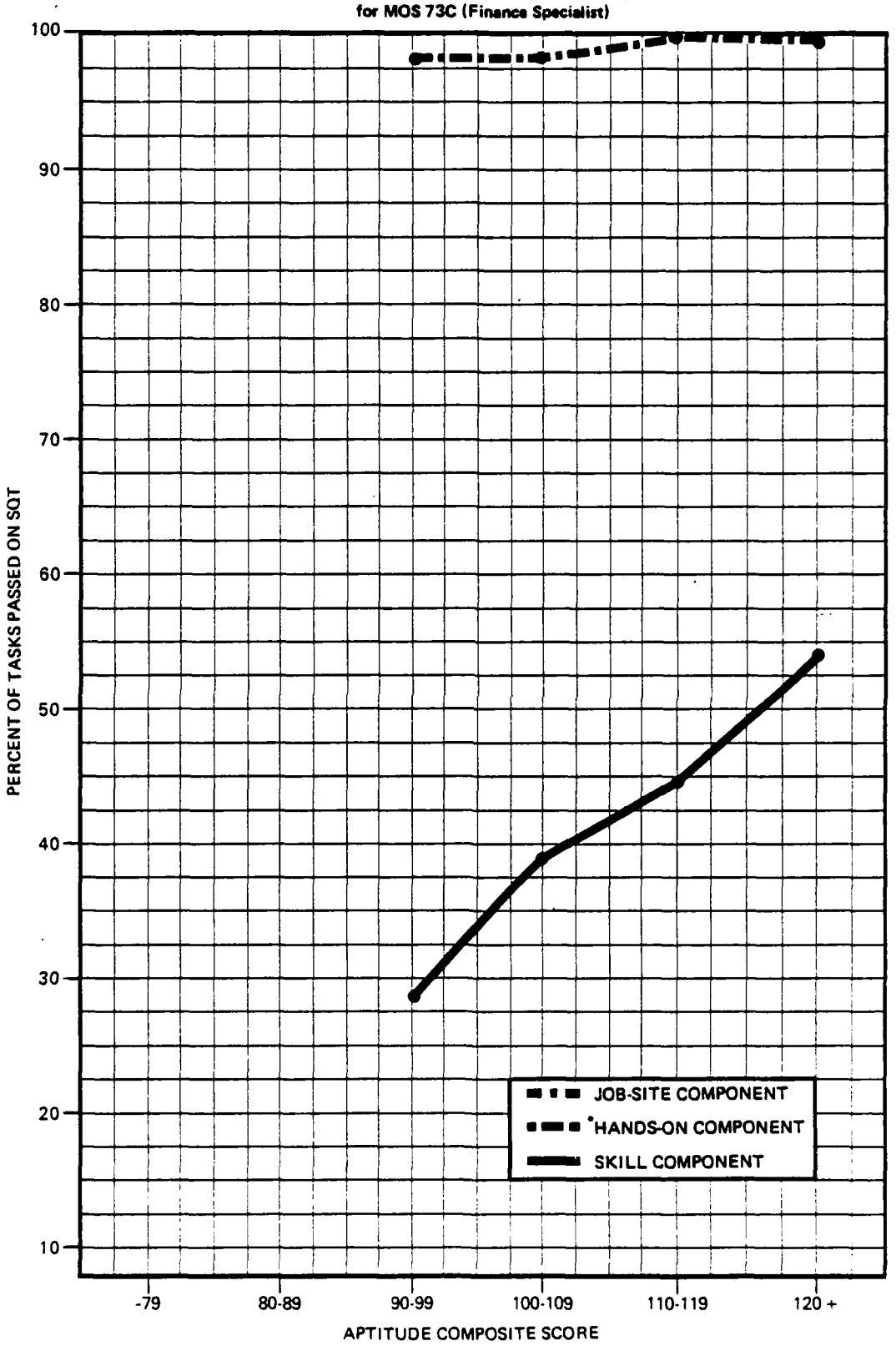


Figure 25
SQT Performance as a Function of Aptitude Composite Score
for MOS 73C (Finance Specialist)



*There was no HOC test for 73C in 1980, when most of the data was obtained.

Figure 26
SQT Performance as a Function of Aptitude Composite Score
for MOS 75B (Personnel Administration Specialist)

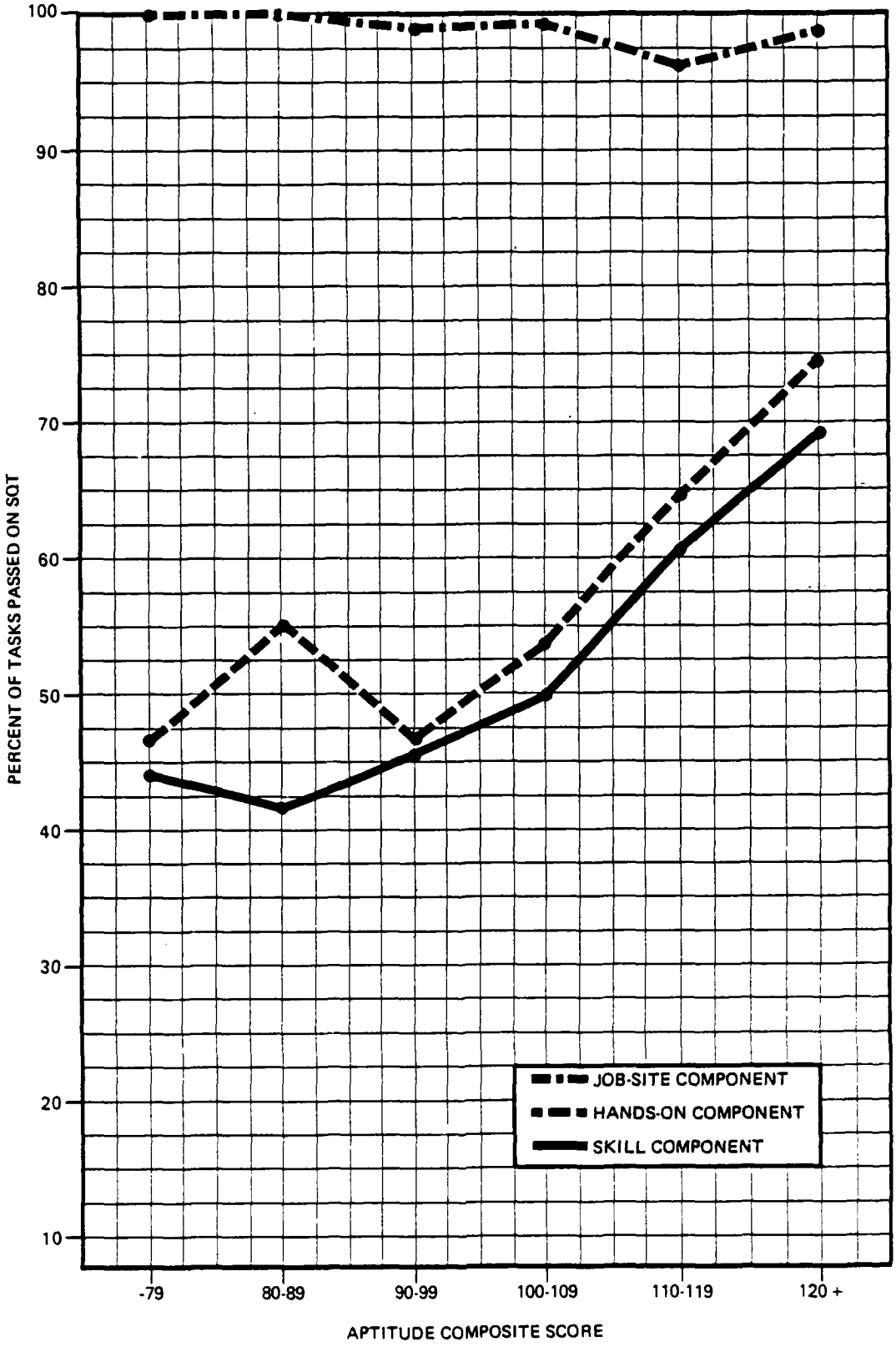


Figure 27
SQT Performance as a Function of Aptitude Composite Score and Education for MOS 11B (Infantryman)

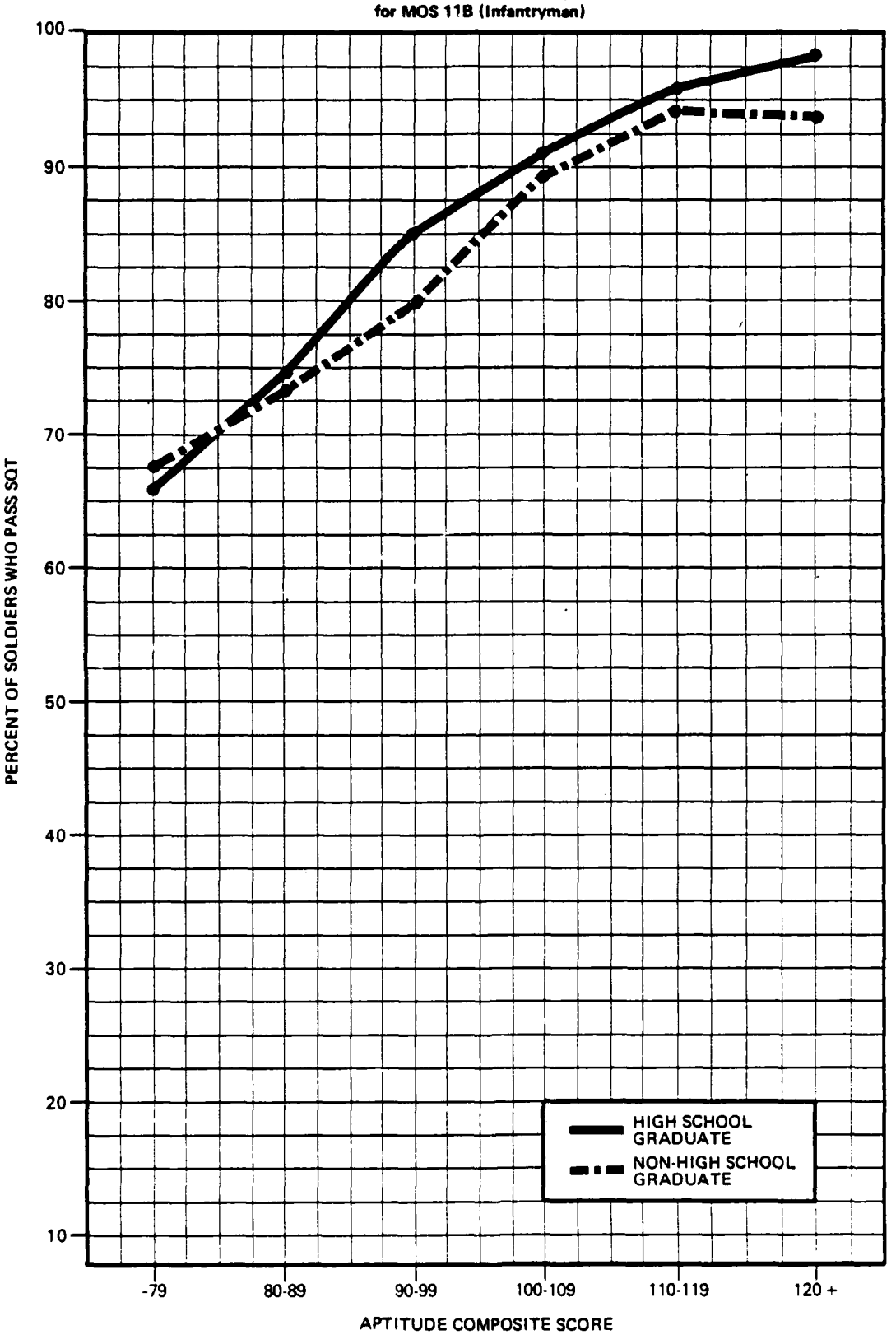


Figure 28
SQT Performance as a Function of Aptitude Composite Score and Education for MOS 11C (Indirect Fire Infantryman)

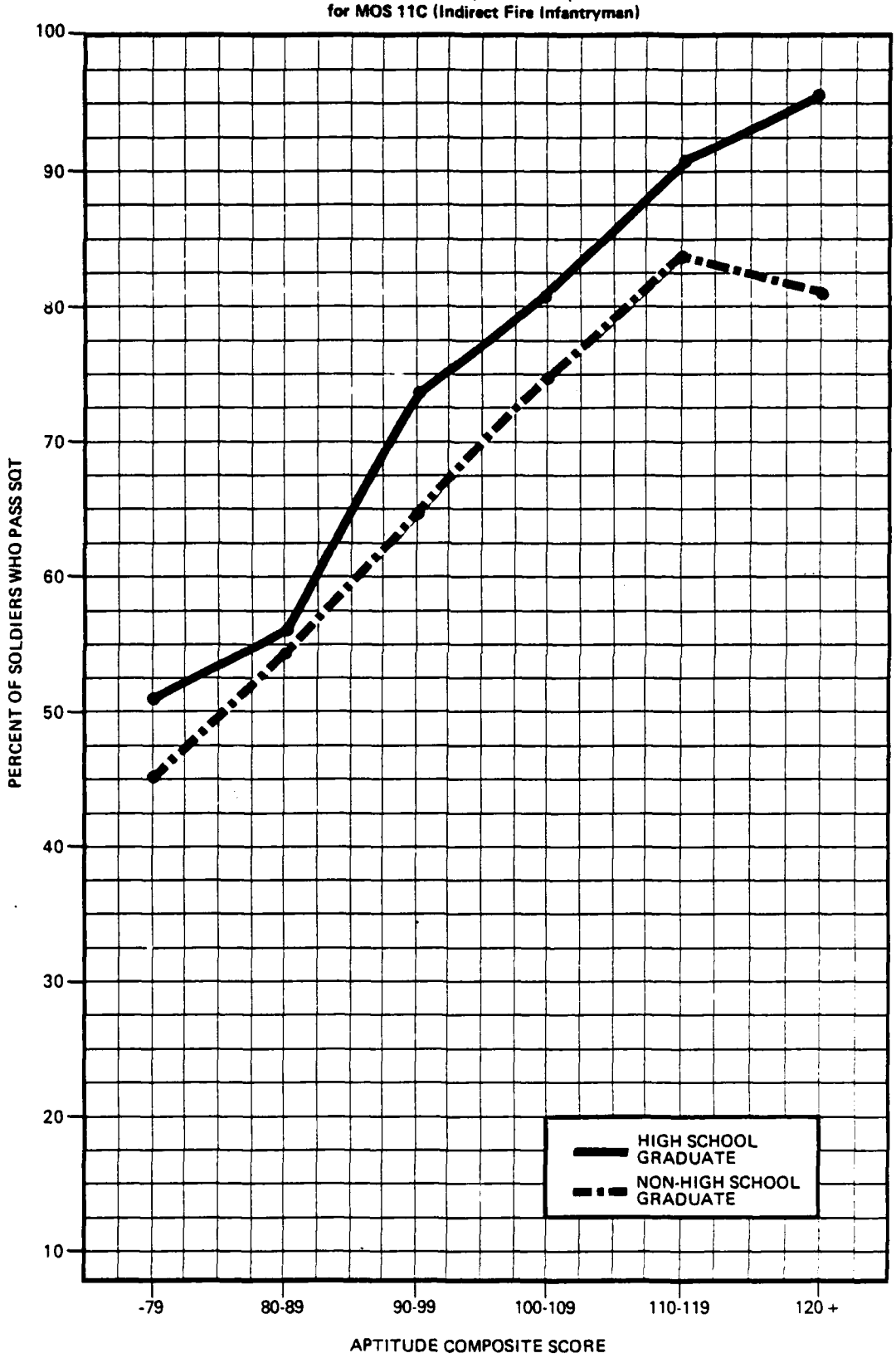


Figure 29
SQT Performance as a Function of Aptitude Composite Score and Education for MOS 19E (Armor Crewman)

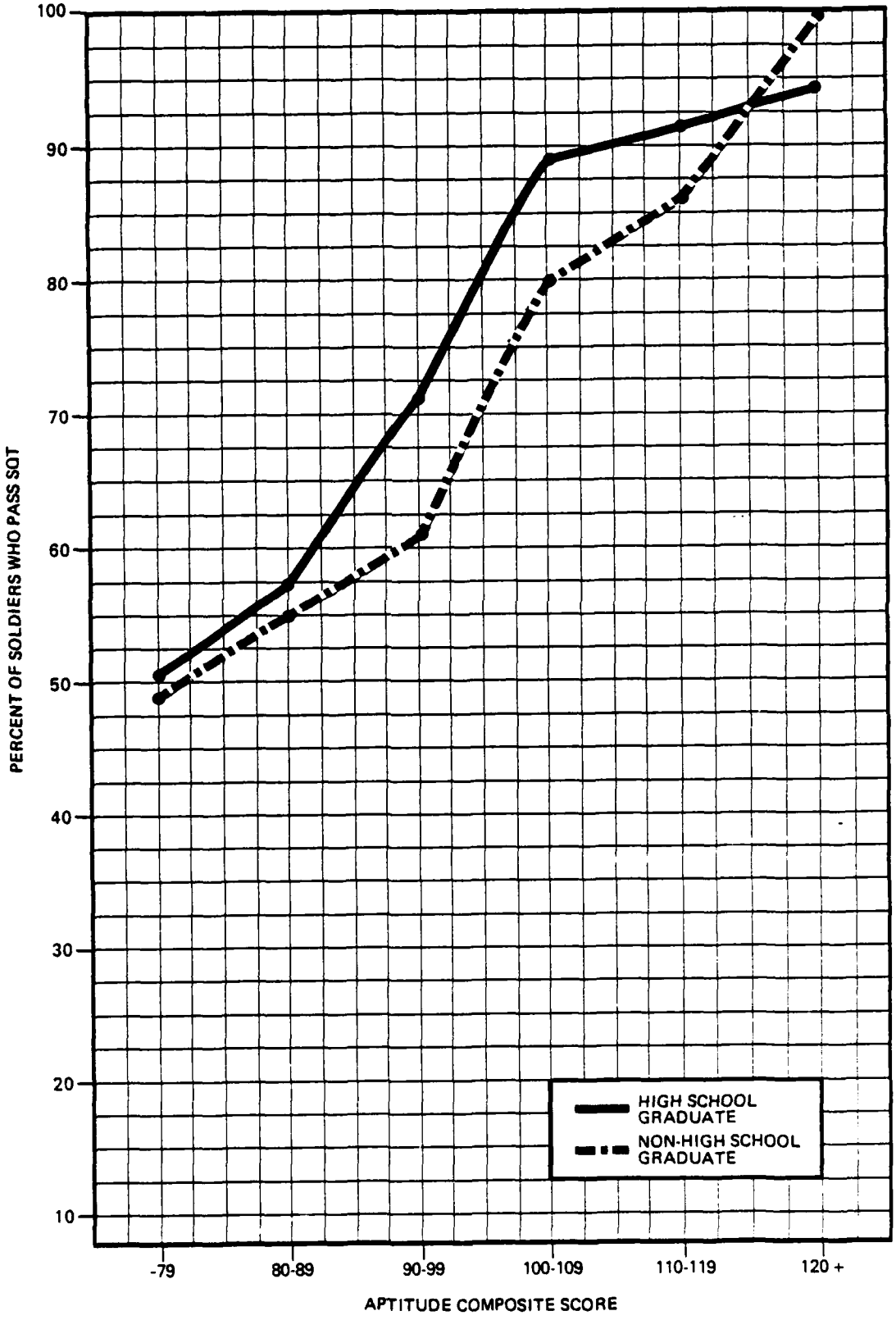


Figure 30
SQT Performance as a Function of Aptitude Composite Score and Education for MOS 05C (Radio Teletypewriter Operator)

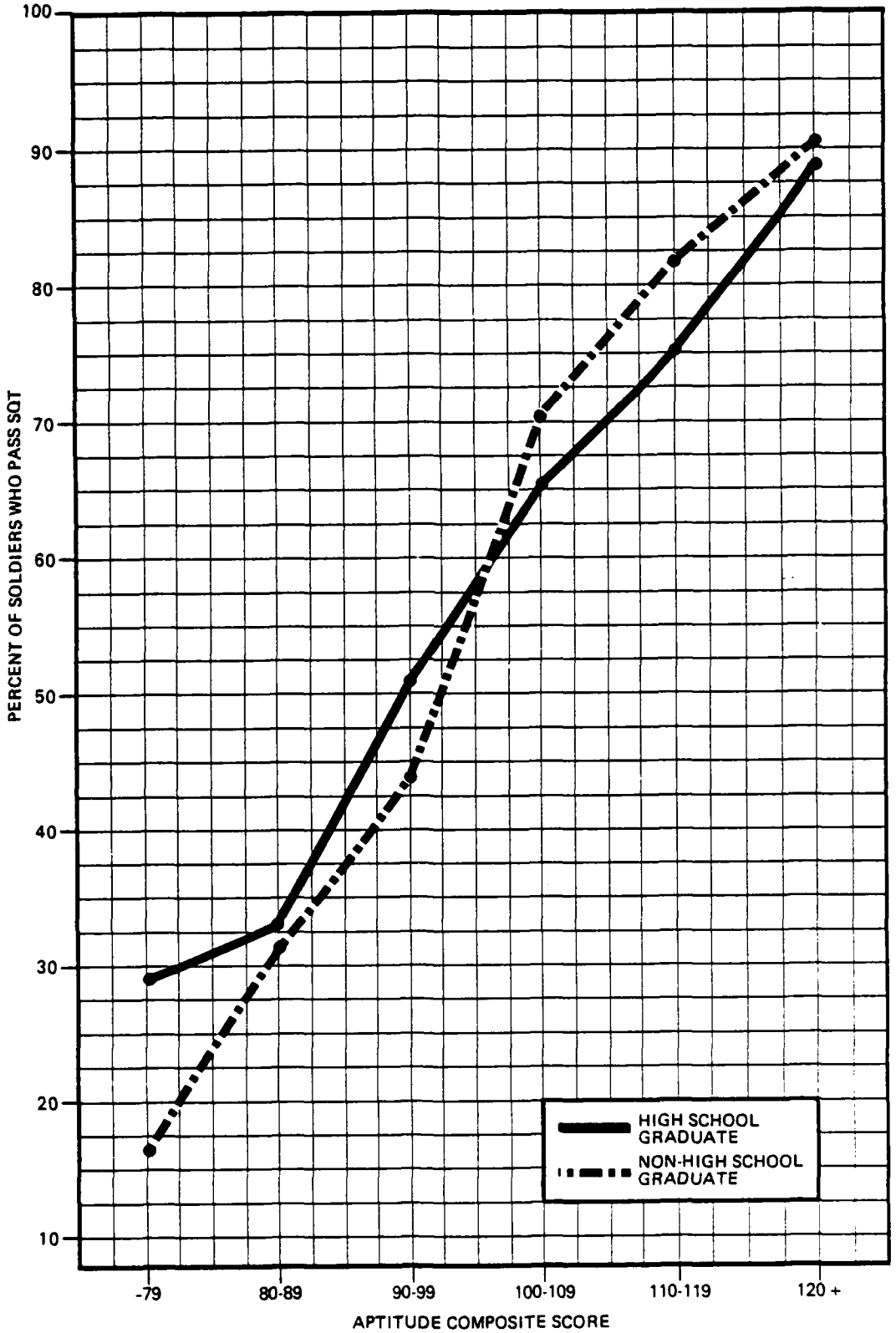


Figure 31
SQT Performance as a Function of Aptitude Composite Score and Education for MOS 31M (Multichannel Communications Operator)

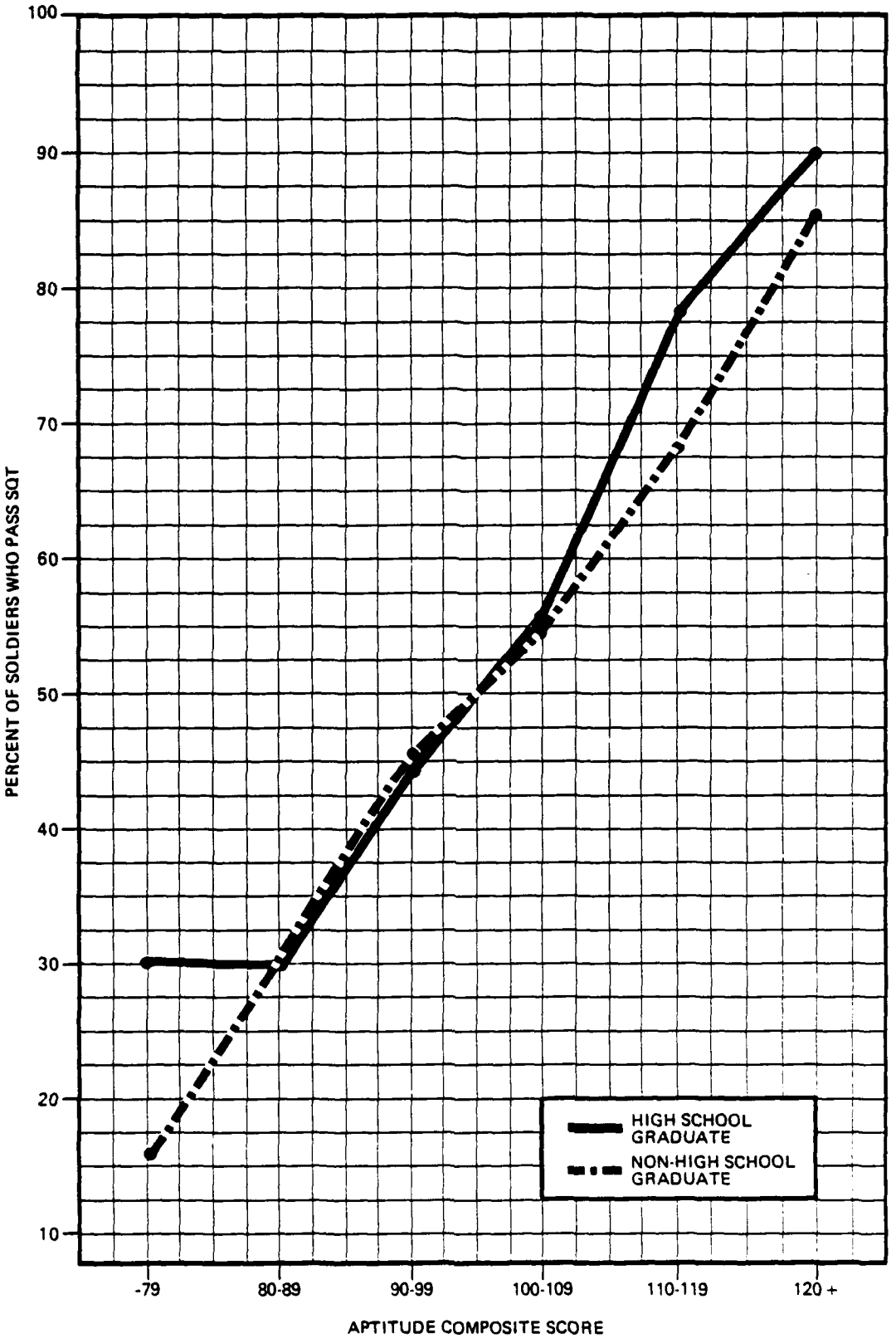


Figure 32
SQT Performance as a Function of Aptitude Composite Score and Education for MOS 67N (Utility Helicopter Repairer)

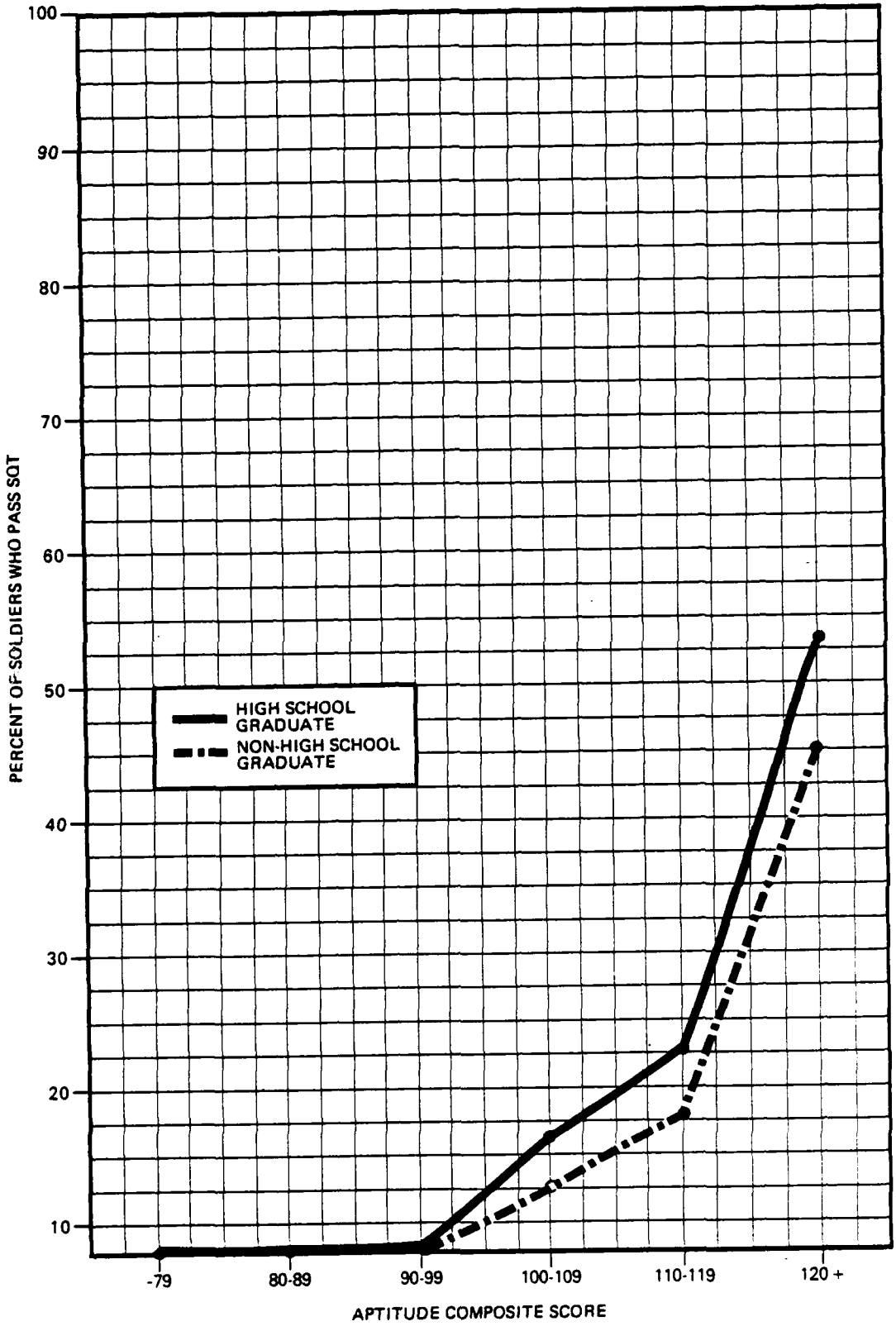


Figure 33
SQT Performance as a Function of Aptitude Composite Score and Education for MOS 73C (Finance Specialist)

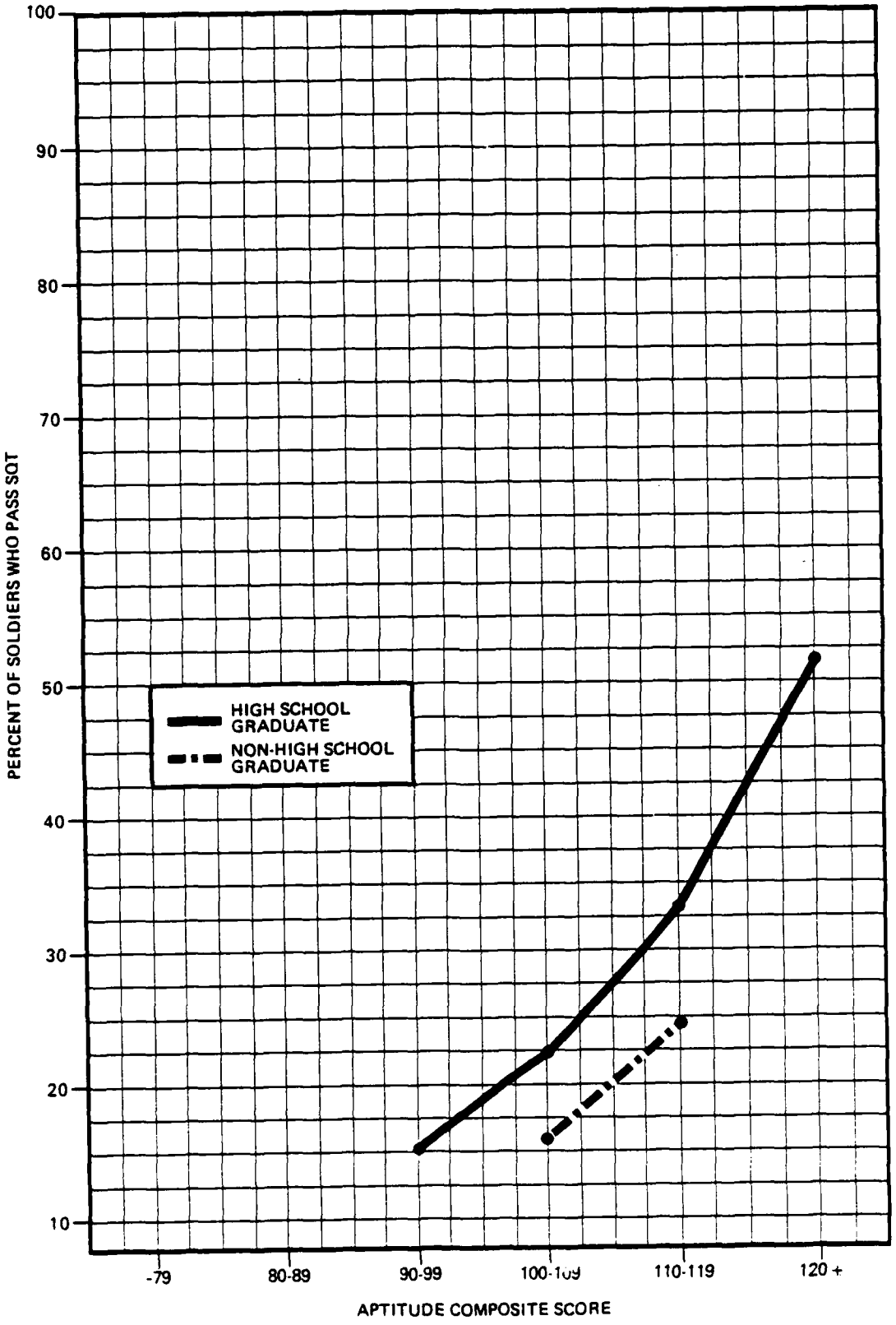


Figure 34
SQT Performance as a Function of Aptitude Composite Score and Education for MOS 75B (Personnel Administration Specialist)

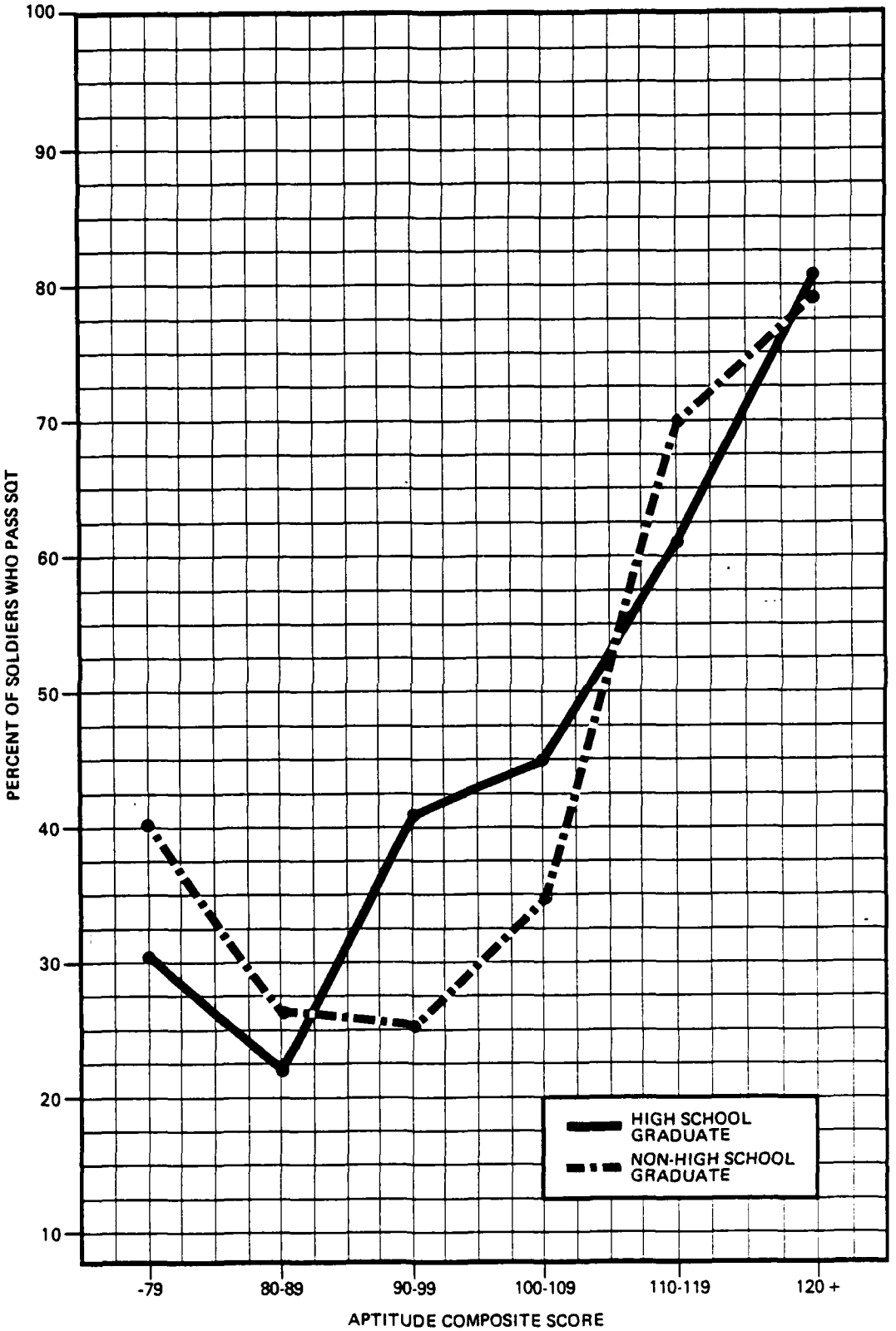


Figure 35
SQT Performance as a Function of Racial/Ethnic Groups and AFQT Category for MOS 11B (Infantryman)

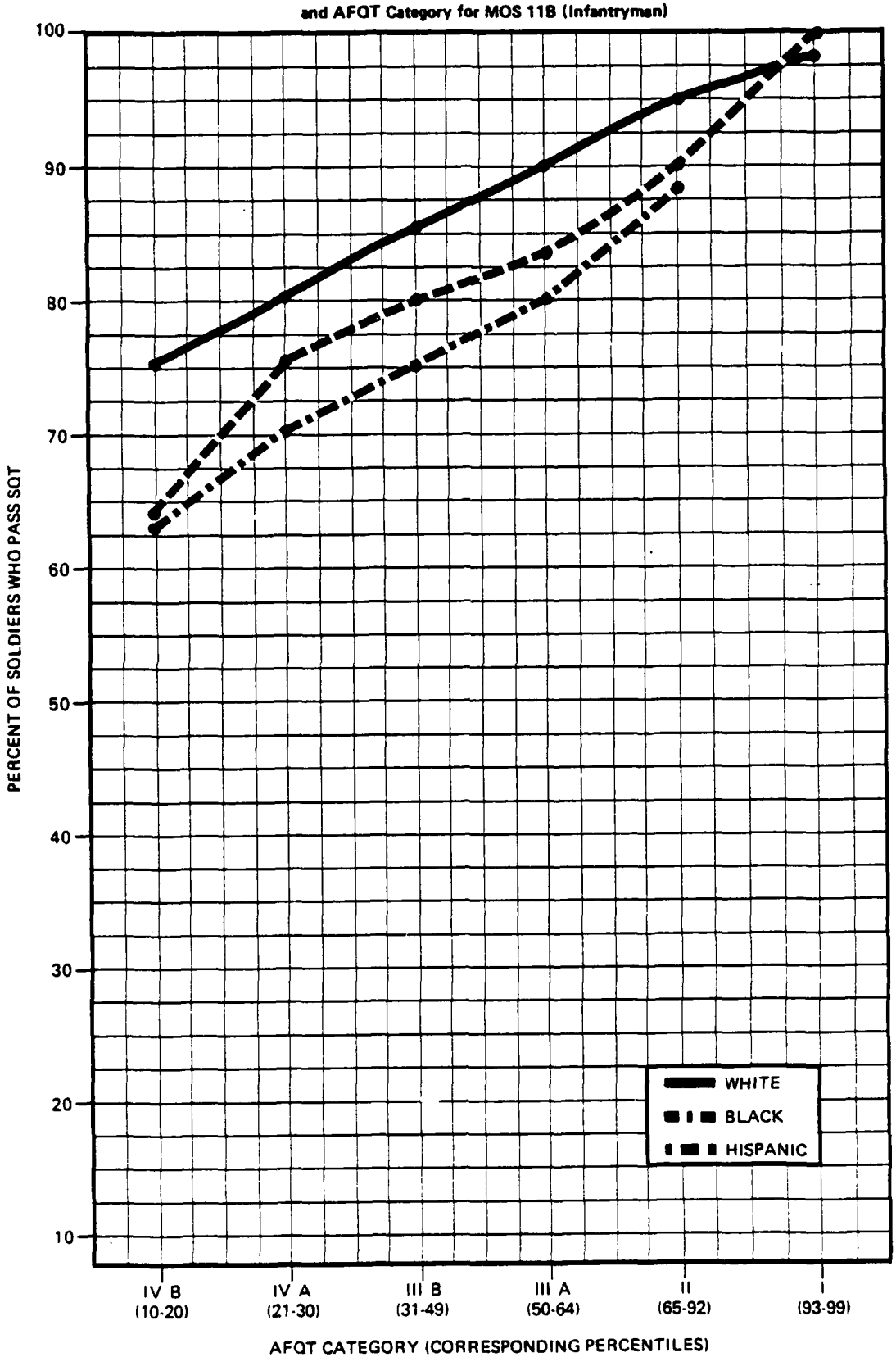


Figure 36
SQT Performance as a Function of Racial/Ethnic Groups
And AFQT Category for MOS 11C (Indirect Fire Infantryman)

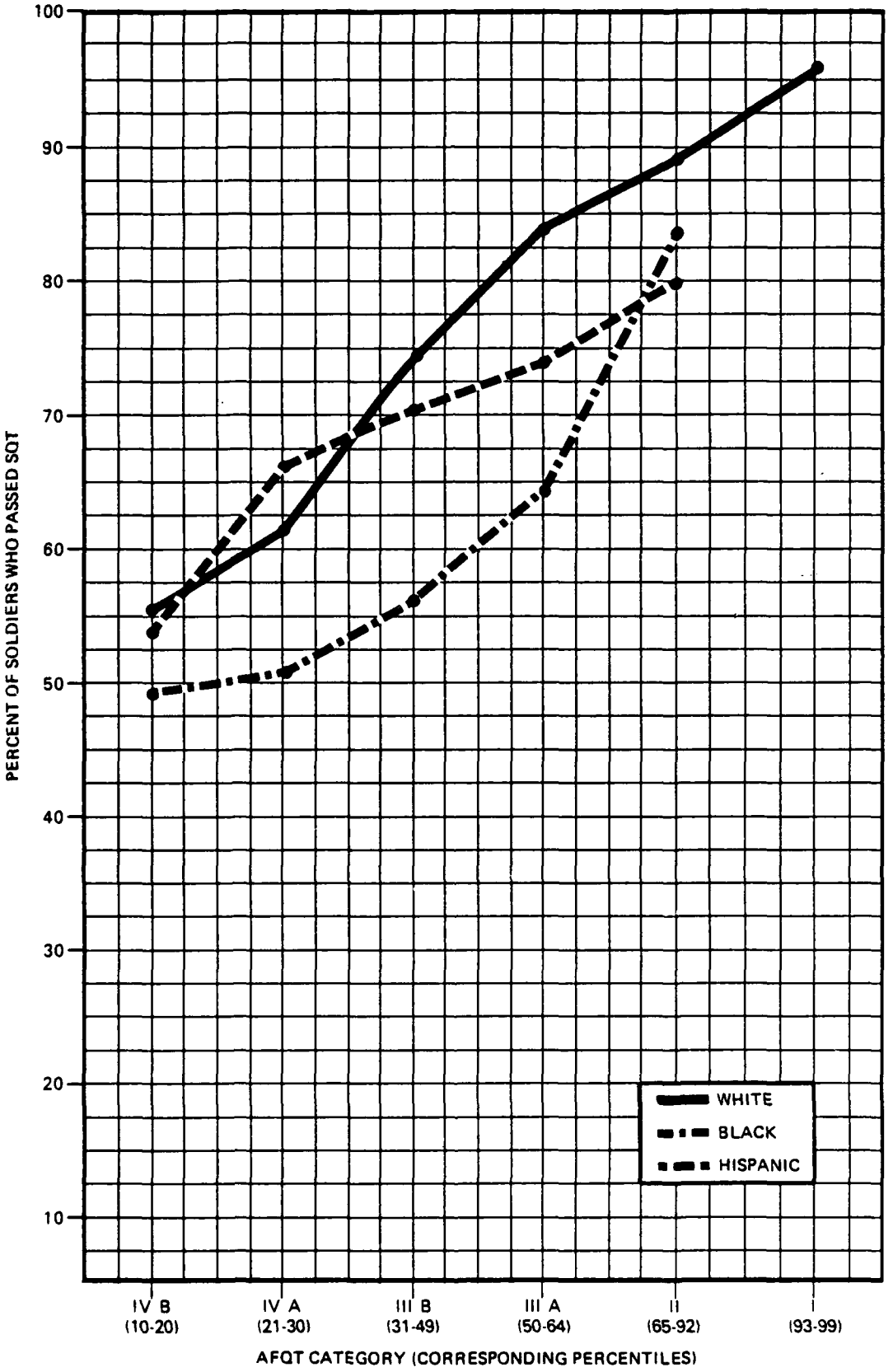


Figure 37
SQT Performance as a Function of Racial/Ethnic Groups
and AFQT Category for MOS 19E (Armor Crewman)

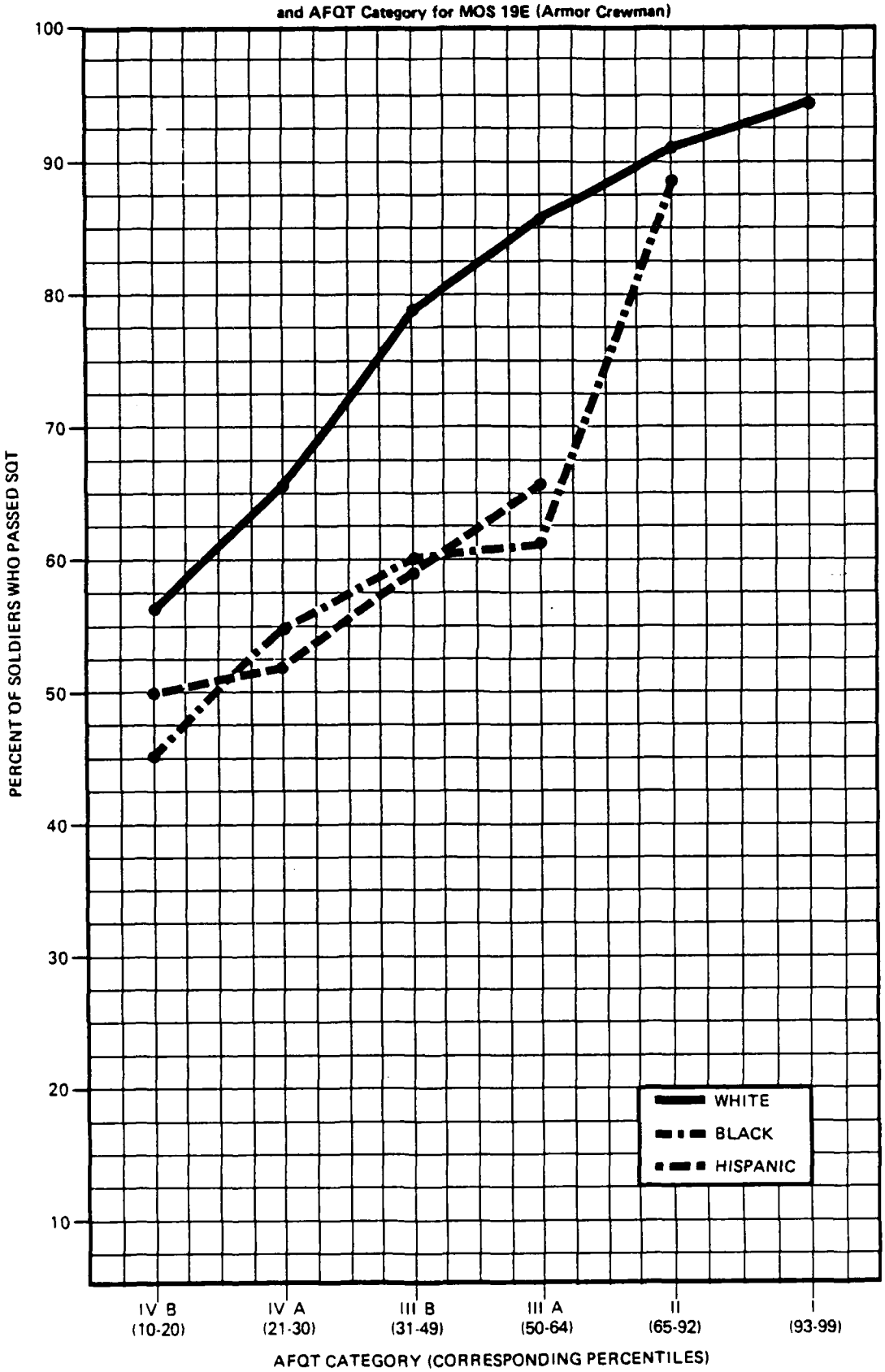


Figure 38
SQT Performance as a Function of Racial/Ethnic Groups
and AFQT Category for MOS 05C (Radio Teletypewriter Operator)

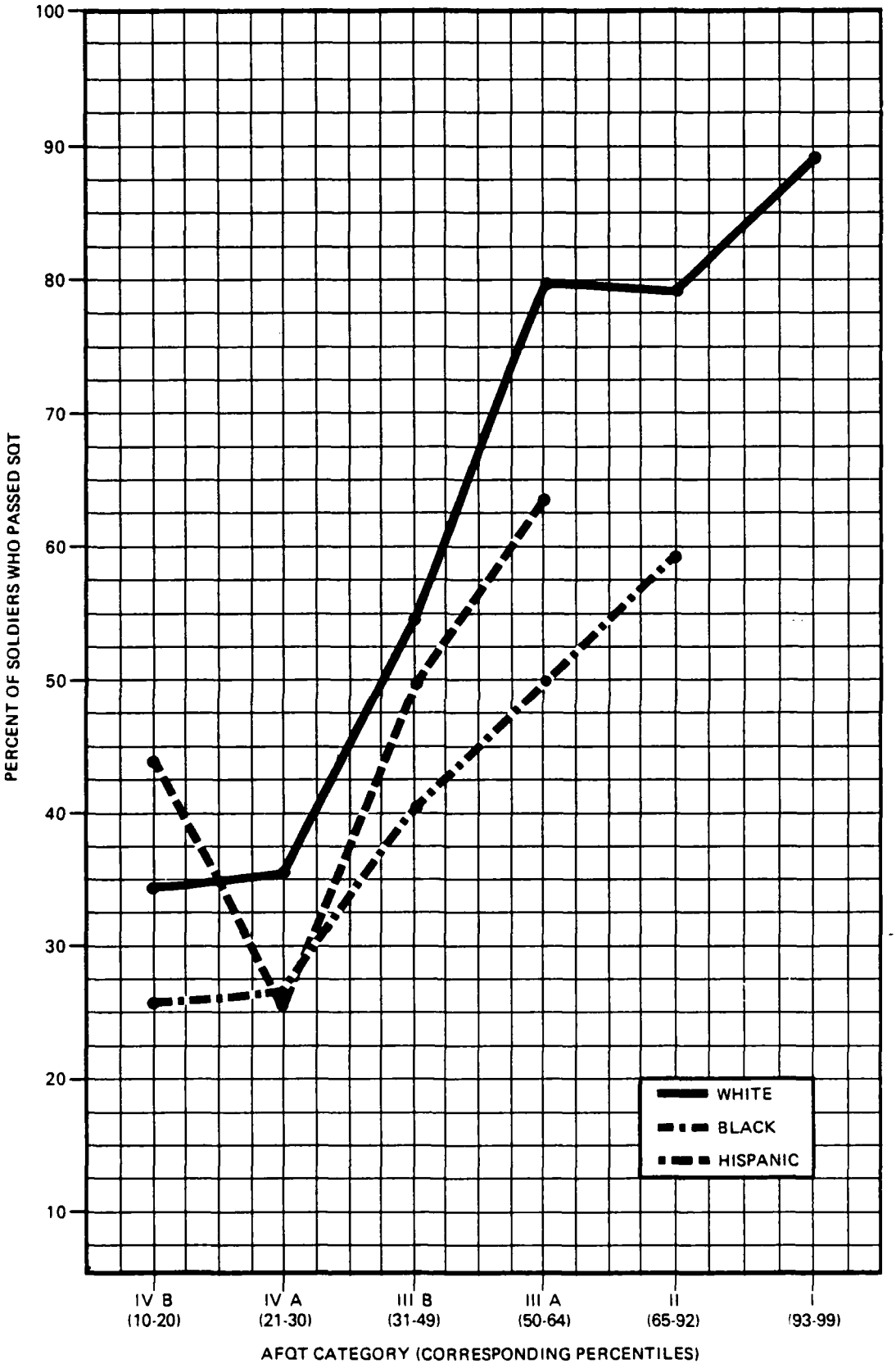


Figure 39
SQT Performance as a Function of Racial/Ethnic Groups and AFQT Category for MOS 31M (Multichannel Communications Operator)

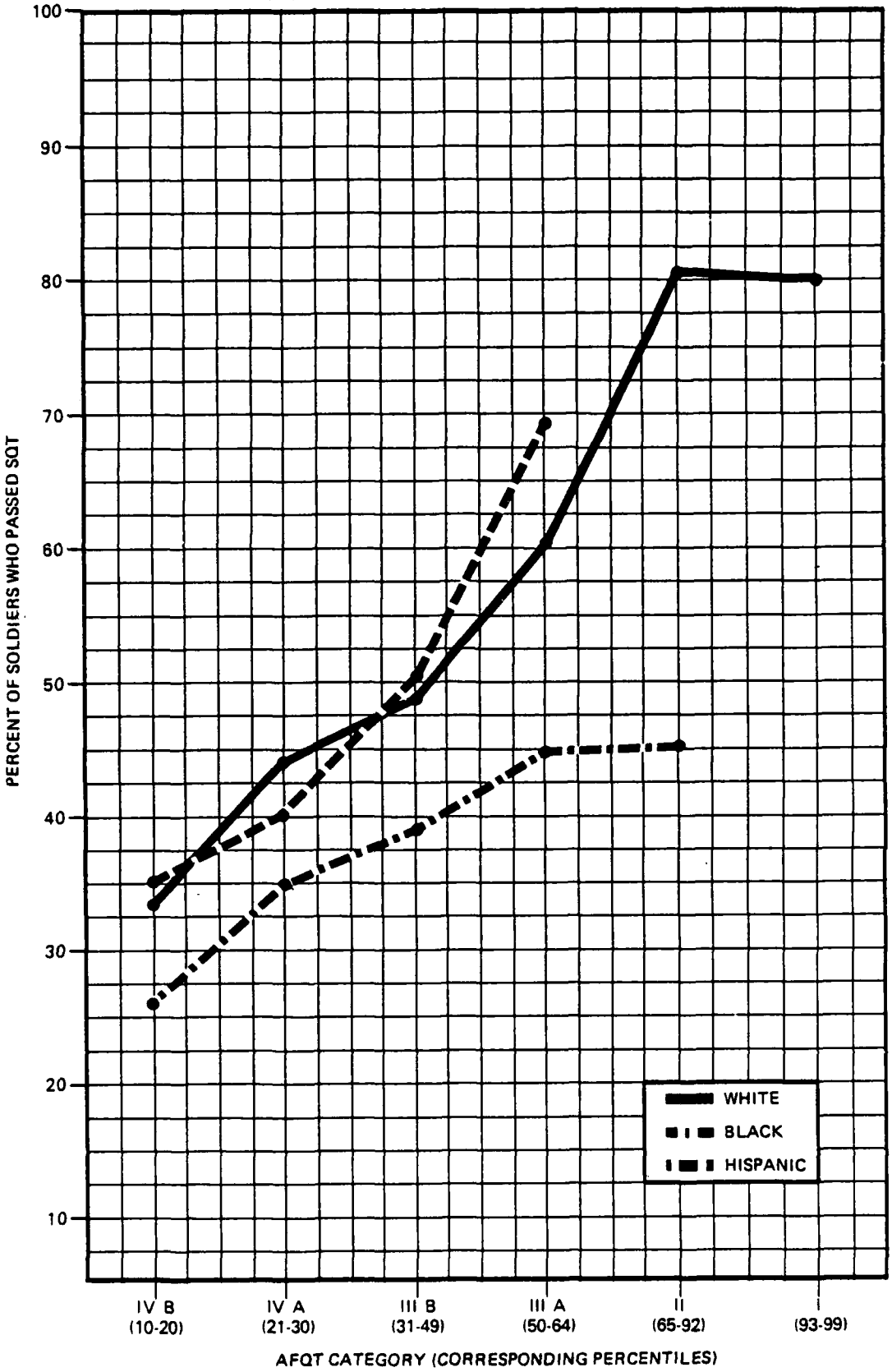


Figure 40
SQT Performance as a Function of Racial/Ethnic Groups and AFQT Category for MOS 67N (Utility Helicopter Repairer)

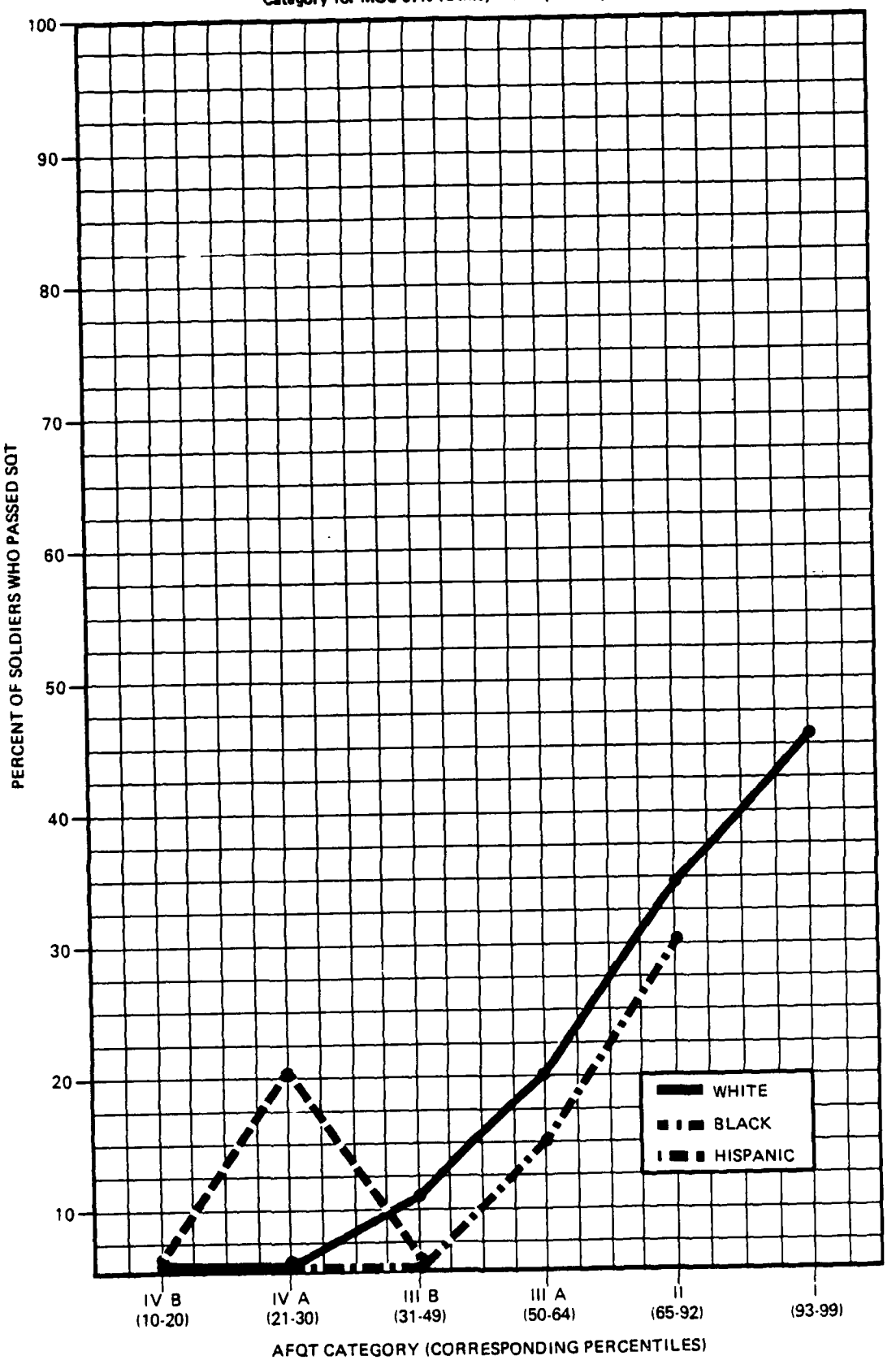


Figure 41
SQT Performance as a Function of Racial/Ethnic Groups and AFQT Category for MOS 73C (Finance Specialist)

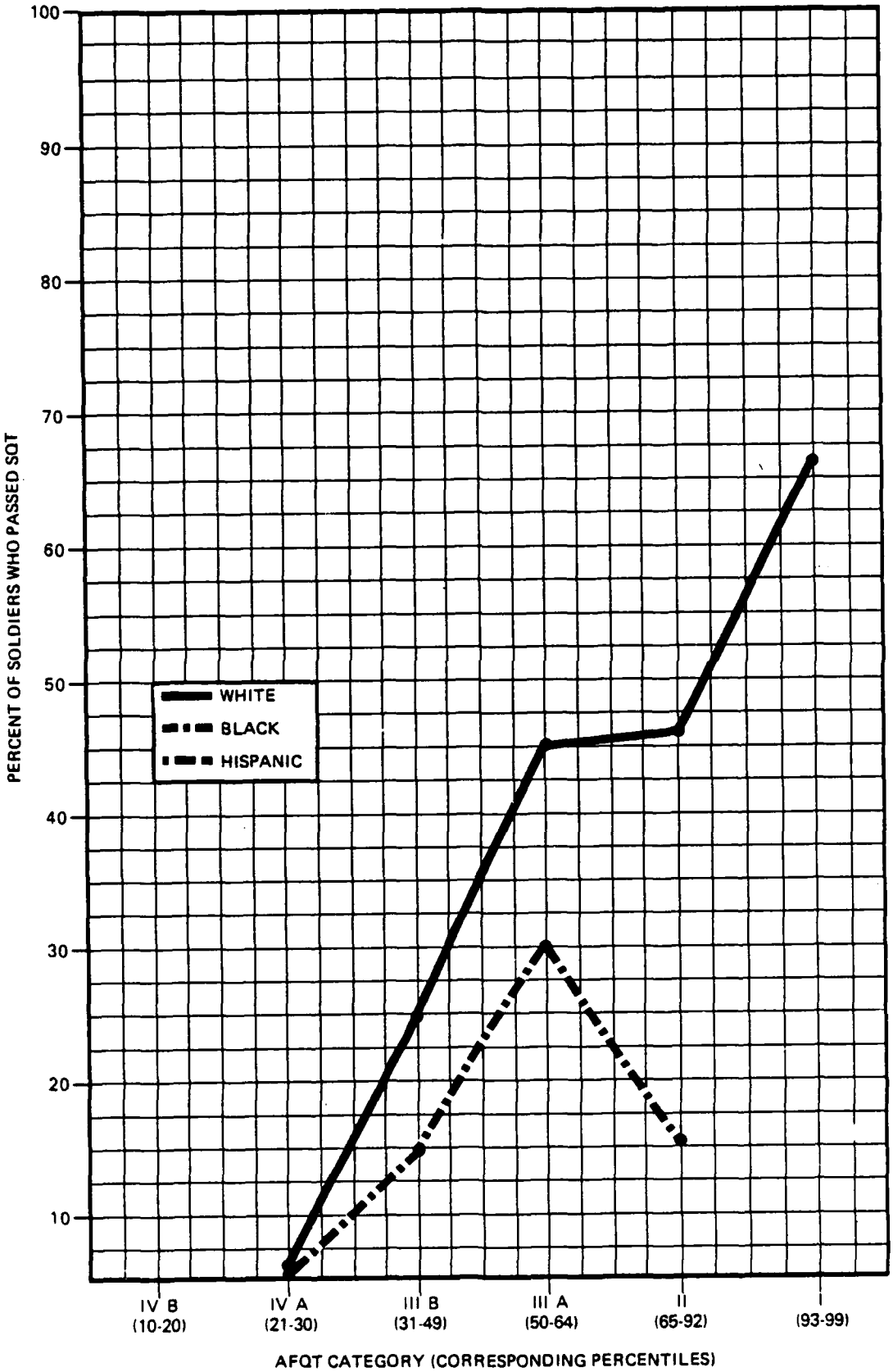


Figure 42
SQT Performance as a Function of Racial/Ethnic Groups and AFQT Category for MOS 75B (Personnel Administration Specialist)

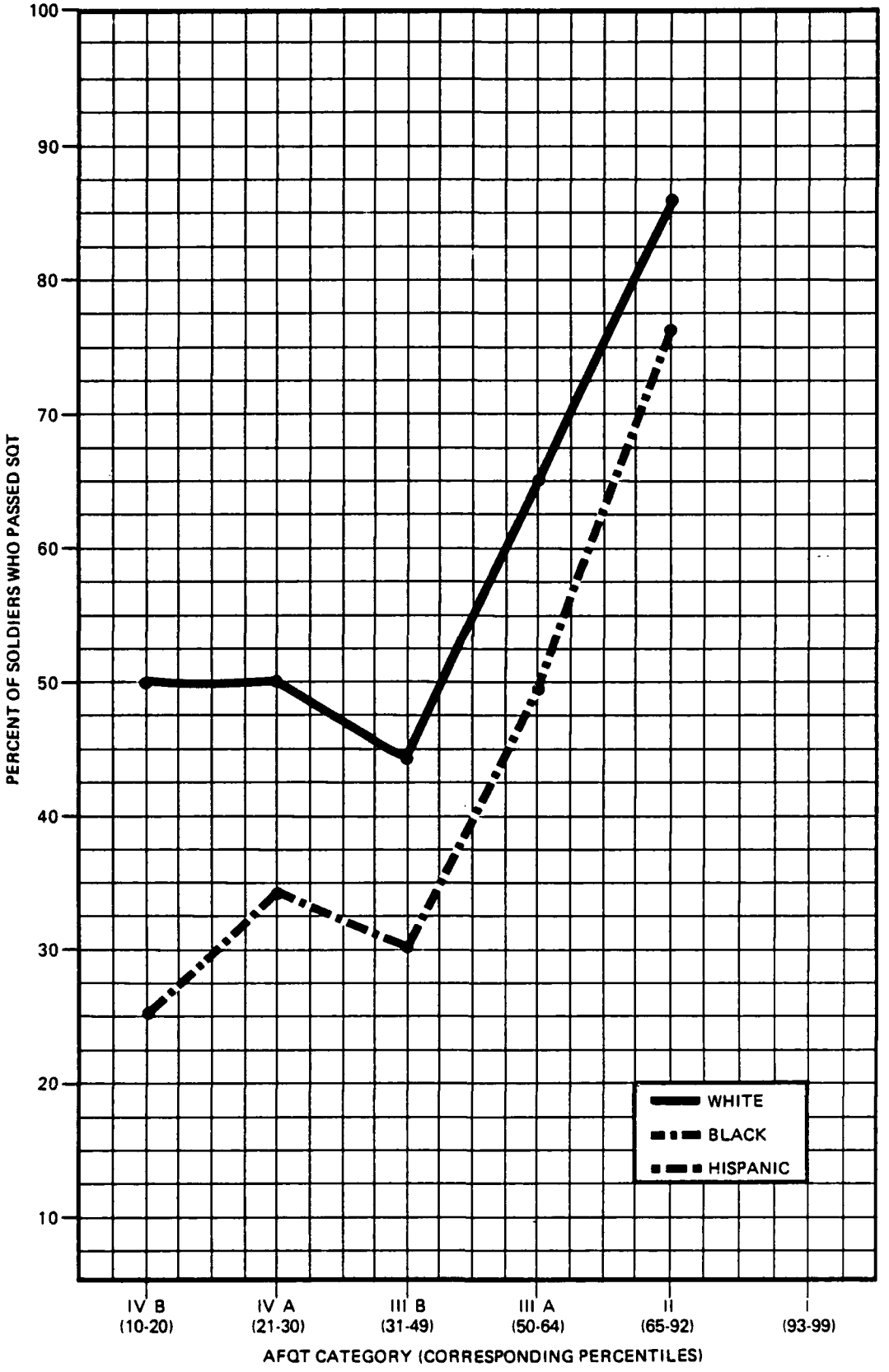


Figure 43
SQT Performance as a Function of Racial/Ethnic Groups and Aptitude Composite Scores for MOS 11B (Infantryman)

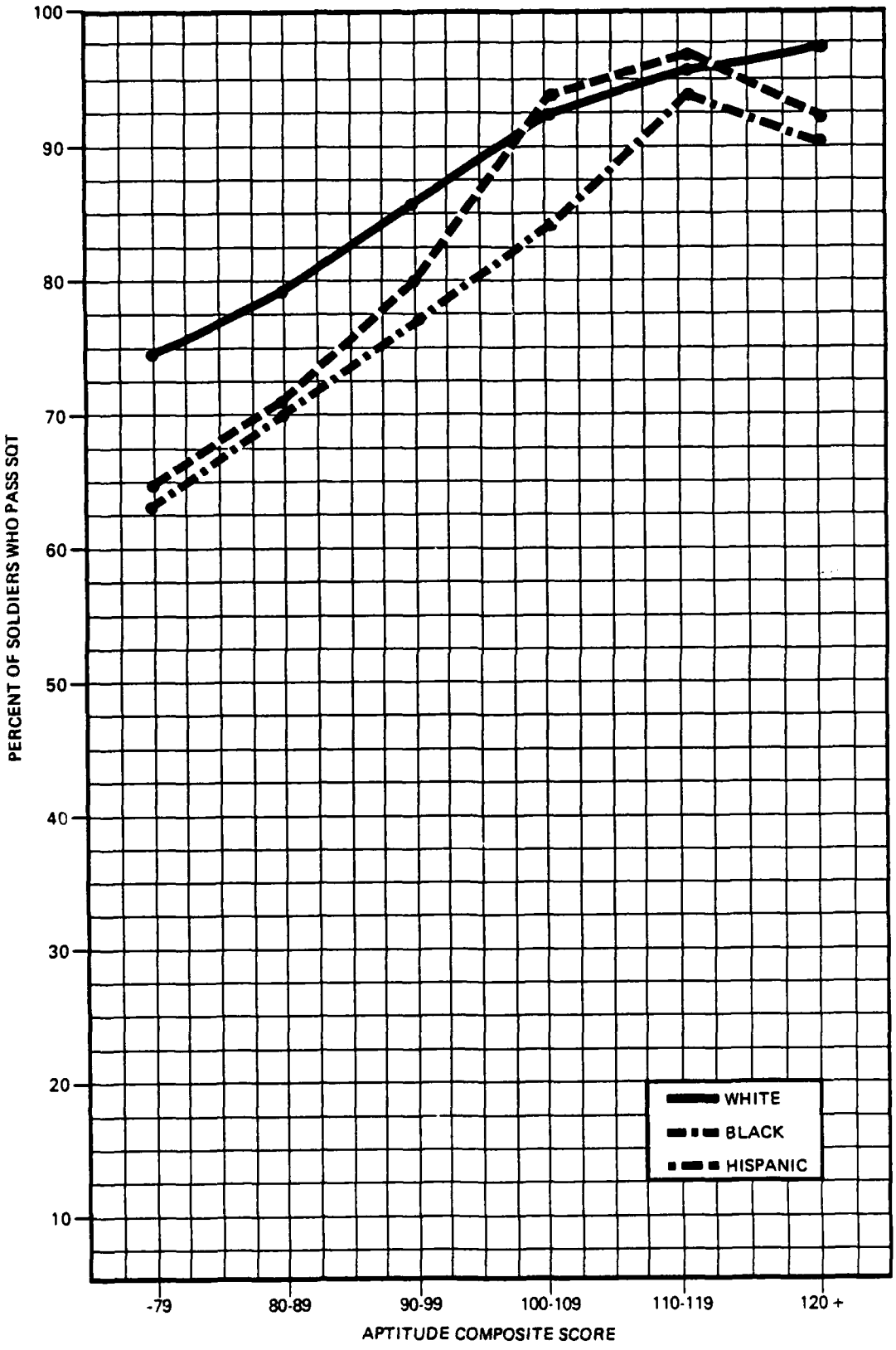


Figure 44
SQT Performance as a Function of Racial/Ethnic Groups and Aptitude Composite Scores for MOS 11C (Indirect Fire Infantryman)

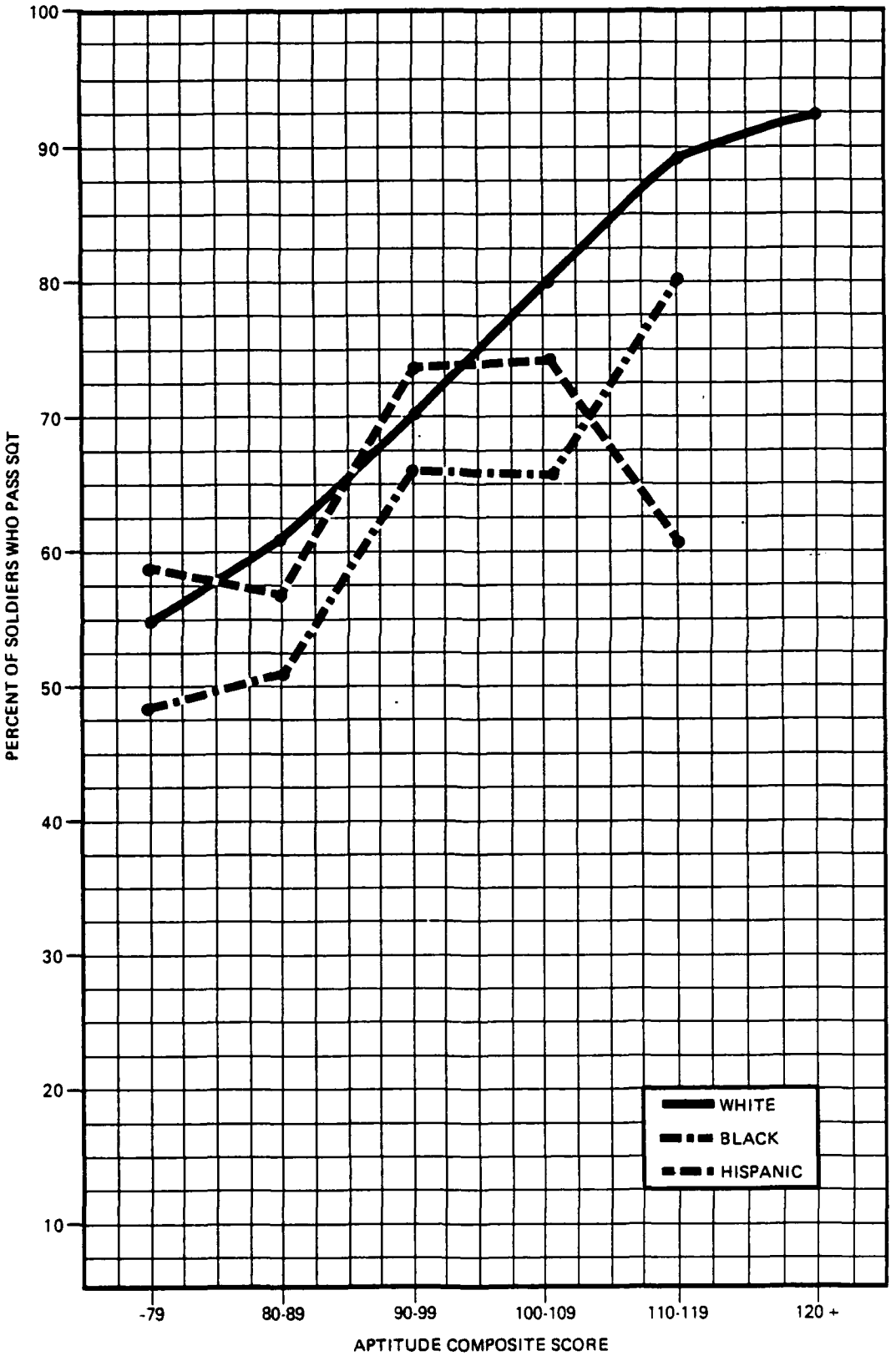


Figure 45
SQT Performance as a Function of Racial/Ethnic Groups and Aptitude Composite Scores for MOS 19E (Armor Crewman)

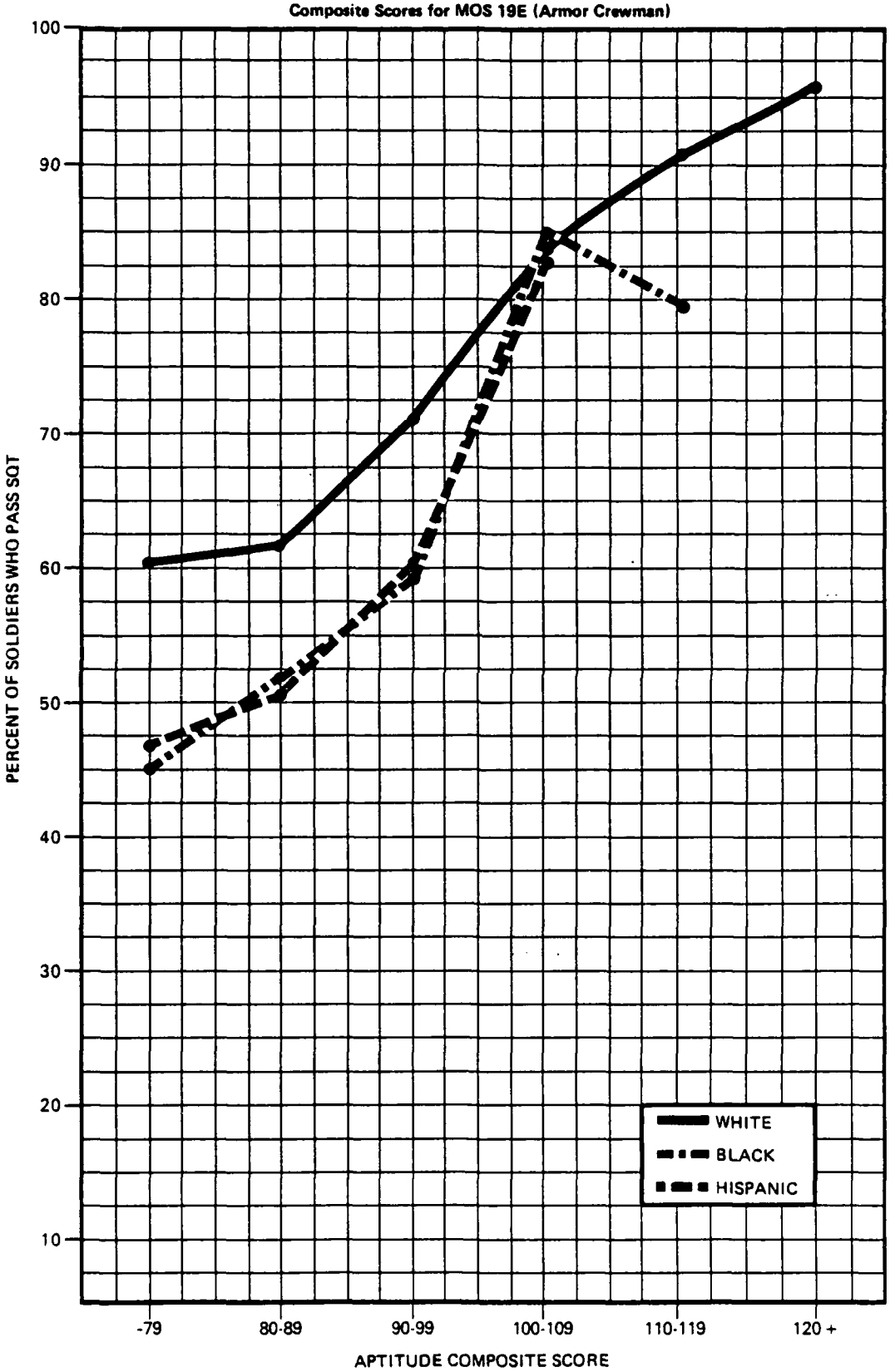


Figure 46
SQT Performance as a Function of Racial/Ethnic Groups and Aptitude Composite Scores for MOS 05C (Radio Teletypewriter Operator)

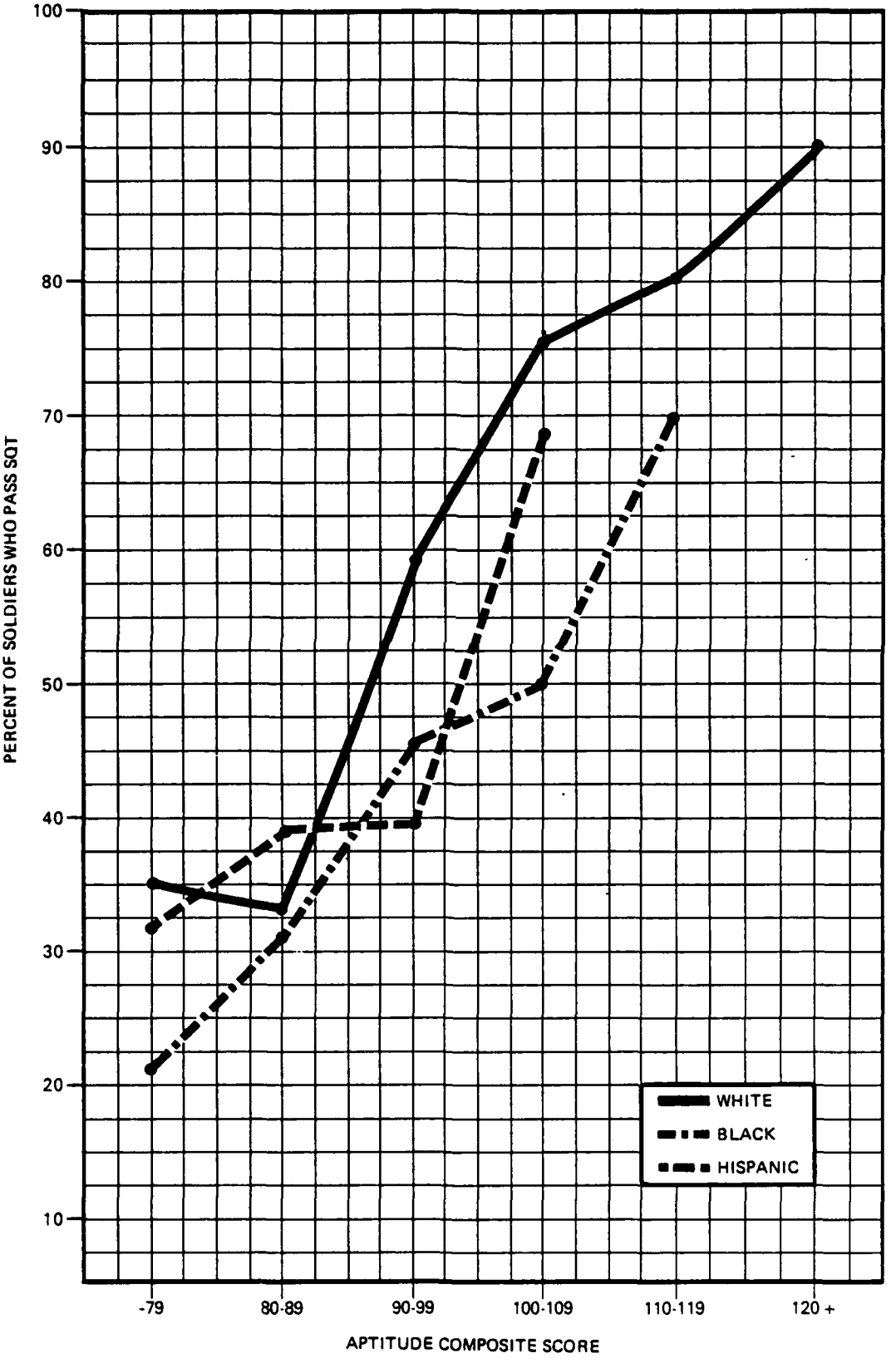


Figure 47
SQT Performance as a Function of Racial/Ethnic Groups and Aptitude Composite Scores for MOS 31M (Multichannel Communications Operator)

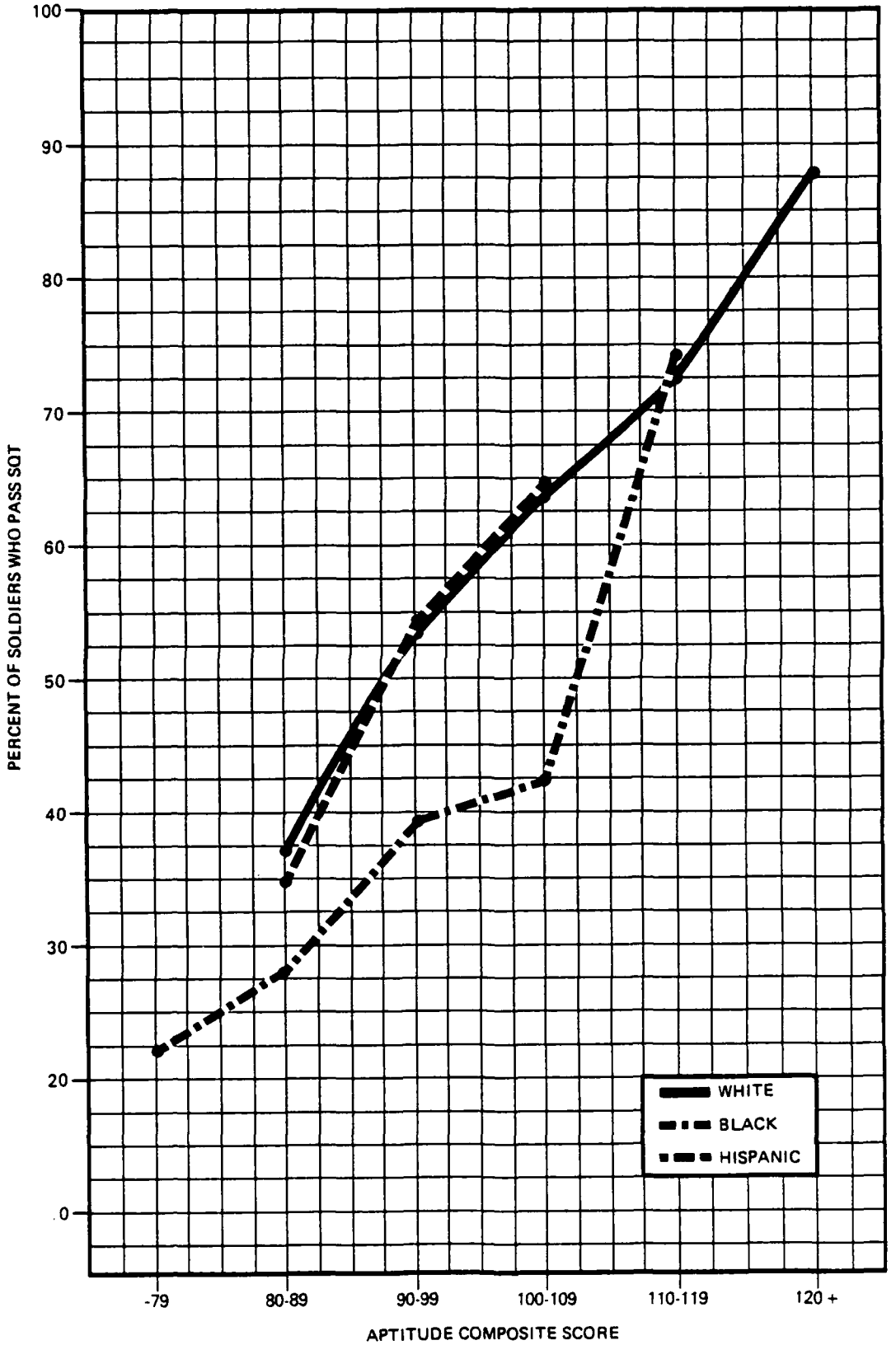


Figure 48
SQT Performance as a Function of Racial/Ethnic Groups and Aptitude Composite Scores for MOS 67N (Utility Helicopter Repairer)

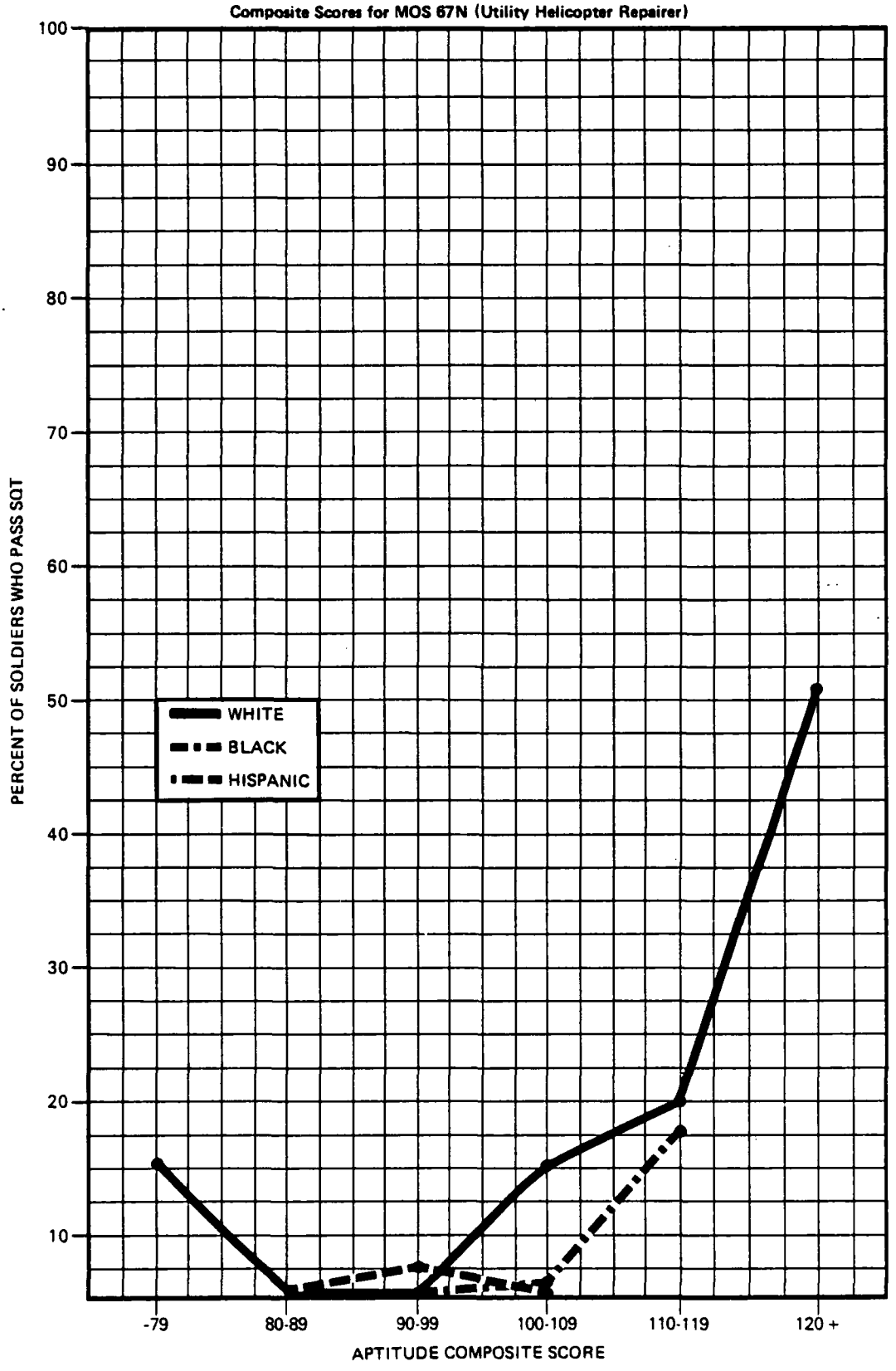


Figure 49
SQT Performance as a Function of Racial/Ethnic Groups and Aptitude Composite Scores for MOS 73C (Finance Specialist)

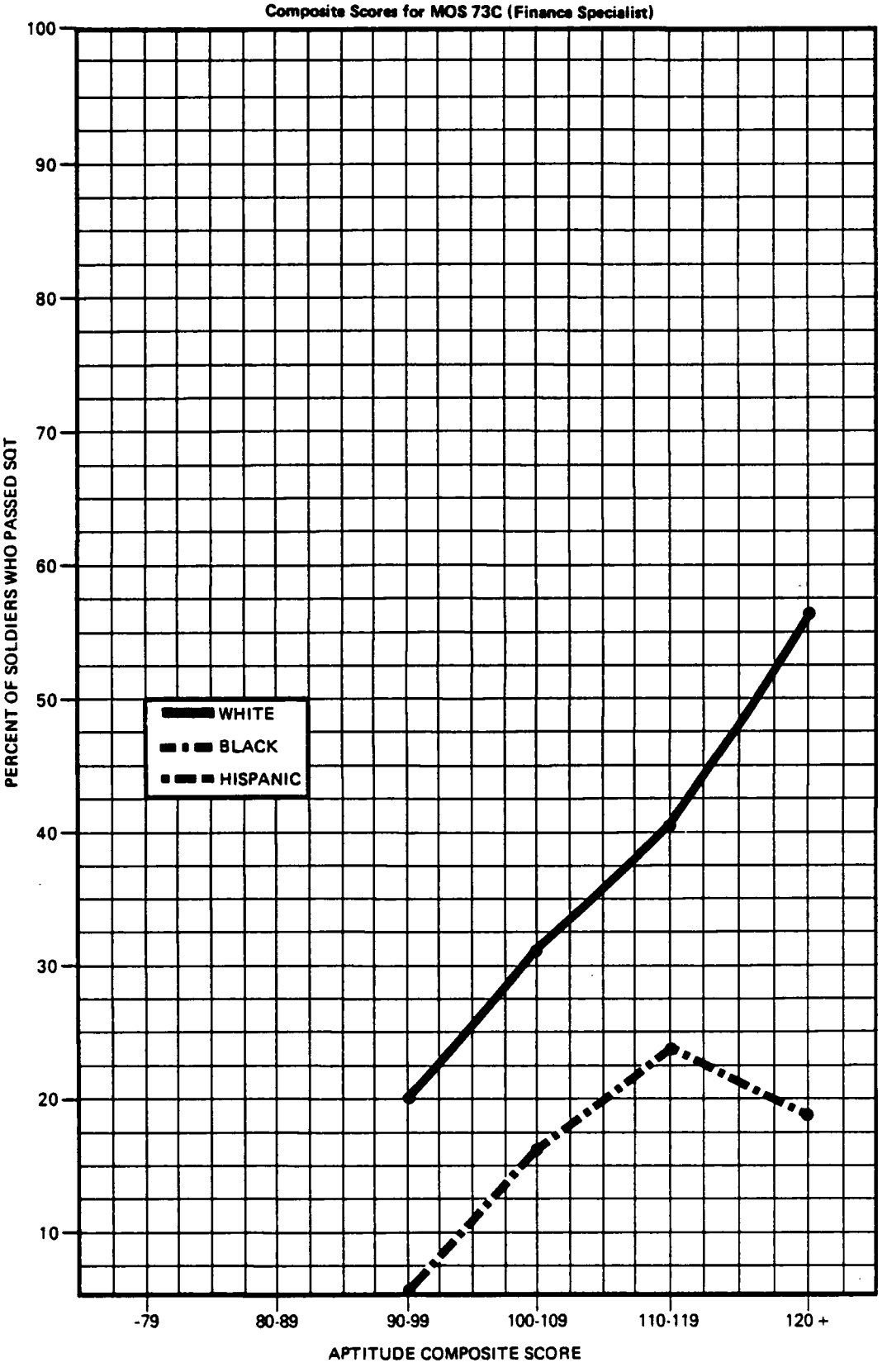
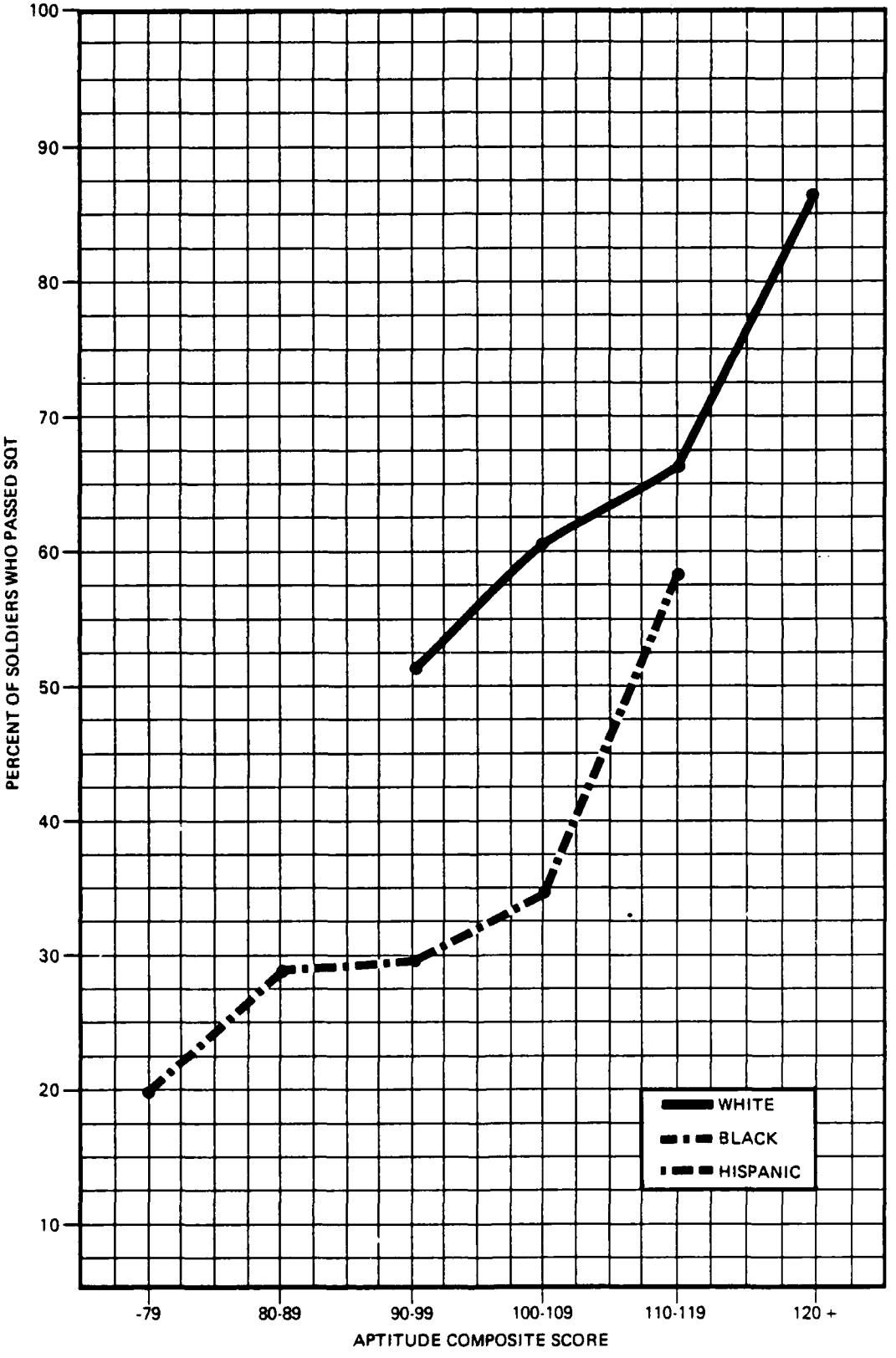


Figure 50
SQT Performance as a Function of Racial/Ethnic Groups and Aptitude Composite Scores for MOS 75B (Personnel Administration Specialist)



Appendix E

UNCORRECTED CORRELATIONS BETWEEN AFQT AND APTITUDE COMPOSITE SCORES AND SQT PERFORMANCE

This Appendix contains Tables 24 through 27, which display the uncorrected correlations between AFQT and aptitude composite scores and SQT performance. The tables correspond to tables 5 through 8 in Section II, where coefficients displayed are corrected for restriction of range.

Table 24

Uncorrected Correlations of AFQT Scores With SQT Performance for Eight Army MOS

		Percent of Tasks Go on:			
MOS	n	Job-Site	Hands-On	Skill	Total SQT
11B	24665	.03	.16	.48	.43
11C	5806	.03	.16	.49	.39
19E	4142	.04	.12	.45	.40
05C	1737	.03	.09	.45	.41
31M	2291	.00	.07	.38	.30
67N	1394	.04	.17	.44	.44
73C	634	.09	--*	.45	.43
75B	477	.01	.23	.42	.44

* No HOC in 1980 SQT for 73C.

Table 25

Uncorrected Correlations of Aptitude Composite Scores with SQT Performance for Eight Army MOS

			Percent of Tasks Go on:			
MOS	Composite	n	Job-Site	Hands-On	Skill	Total SQT
11B	CO	24665	.03	.18	.45	.42
11C	CO	5806	.02	.19	.46	.39
19E	CO	4142	.05	.12	.44	.41
05C	SC	1737	.03	.09	.48	.43
31M	EL	2291	.00	.10	.36	.31
67N	MM	1394	.16	.21	.45	.46
73C	CL	634	.07	--*	.37	.35
75B	CL	477	-.03	.21	.32	.35

* No HOC in 1980 SQT for 73C.

Table 26

Uncorrected Correlations Between AFQT Scores and Performance on Skill and Hands-On Components of the SQT for Different Racial/Ethnic Groups

		Percent of Tasks Go on:					
		Skill Component			Hands-On Component		
	MOS	White	Black	Hispanic	White	Black	Hispanic
11B	Infantryman	.47	.25	.36	.16	.09	.09
11C	Indirect Fire Infantryman	.52	.20	.34	.18	.05	.11
19E	Armor Crewman	.45	.22	.19	.13	.07	.03
05C	Radio Teletypewriter Operator	.46	.26	.20	.11	.05	-.05
31M	Multichannel Communications Operator	.46	.18	.34	.11	.01	.20
67N	Utility Helicopter Repairer	.40	.44	.25	.16	.15	.02
73C	Finance Specialist	.35	.26	.40	-.06	.16	.00*
75B	Personnel Administration Specialist	.47	.29	.21	.22	.11	.41

*Too few observations.

Table 27

Uncorrected Correlations Between Aptitude Composite Scores and Performance on Skill and Hands-On Components of the SQT for Different Racial/Ethnic Groups

			Percent of Tasks Go on:					
			Skill Component			Hands-On Component		
	MOS	Aptitude Composite	White	Black	Hispanic	White	Black	Hispanic
11B	Infantryman	CO	.40	.24	.37	.18	.10	.12
11C	Indirect Fire Infantryman	CO	.45	.18	.30	.18	.12	.13
19E	Armor Crewman	CO	.40	.21	.27	.14	.09	.00
05C	Radio Teletypewriter Operator	SC	.48	.29	.25	.11	.05	-.06
31M	Multichannel Communications Operator	EL	.41	.21	.26	.12	.08	.03
67N	Utility Helicopter Repairer	MM	.41	.43	.16	.19	.17	.15
73C	Finance Specialist	CL	.25	.27	.42	.12	-.06	.00*
75B	Personnel Administration Specialist	CL	.39	.18	.23	.24	.12	.33

*Too few observations.

Appendix F

RELATIONSHIP BETWEEN AFQT/APTITUDE COMPOSITE SCORES AND TRAINING PERFORMANCE MEASURES

Figures 51 through 98, representing the relationship between AFQT/aptitude composite scores and various measures of training performance measures, are contained in Appendix F as follows:

- Figures 51 through 74 look at final course grades as a function of AFQT/ASVAB category and level of education;
- Figures 75 through 82 show final course grades across the AFQT/aptitude composite categories by racial/ethnic group;
- In Figures 83 and 84, attrition in training is plotted for four occupational specialties by AFQT/aptitude composite categories;
- Figures 85 and 86 look at differences in time to complete training across AFQT/aptitude composite categories for four occupational specialties;
- Figures 87 and 88 show the relationship of Mortar Qualification Test scores (in MOS 11C) to AFQT/aptitude composite scores;
- In Figures 89 through 96, peer nomination ratings are plotted as a function of AFQT/aptitude composite category and education for four occupational specialties ; and
- Figures 97 and 98 show the relationship between instructor ratings and AFQT/aptitude composite scores as a function of level of education for MOS 11B (Infantryman).

Figure 51
Final Course Grade as a Function of AFQT Category and Level of Education for Marine Corps Specialty 0311 (Infantryman)

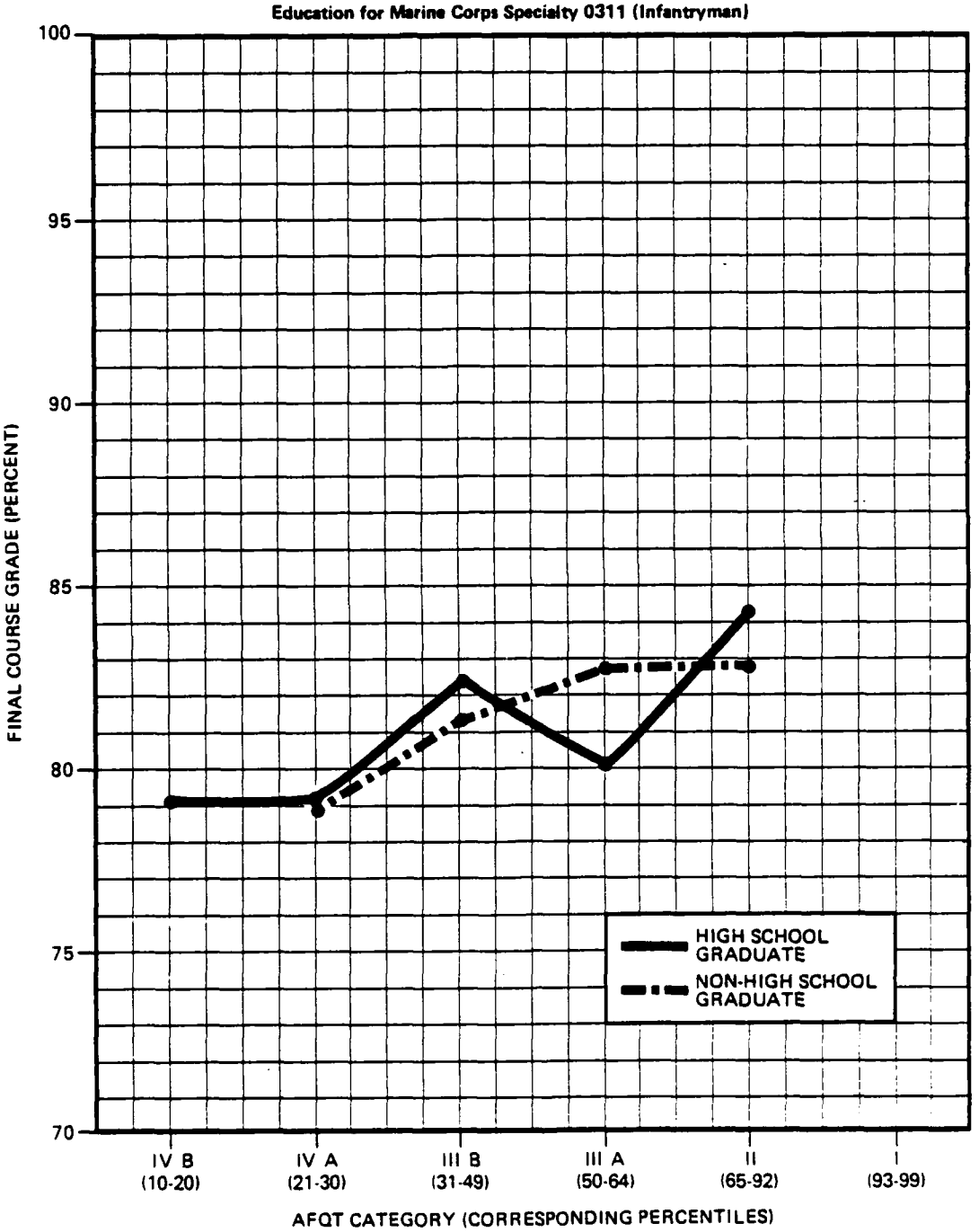


Figure 52
Final Course Grade as a Function of AFQT Category and Level of Education for Army MOS 11B (Infantryman)

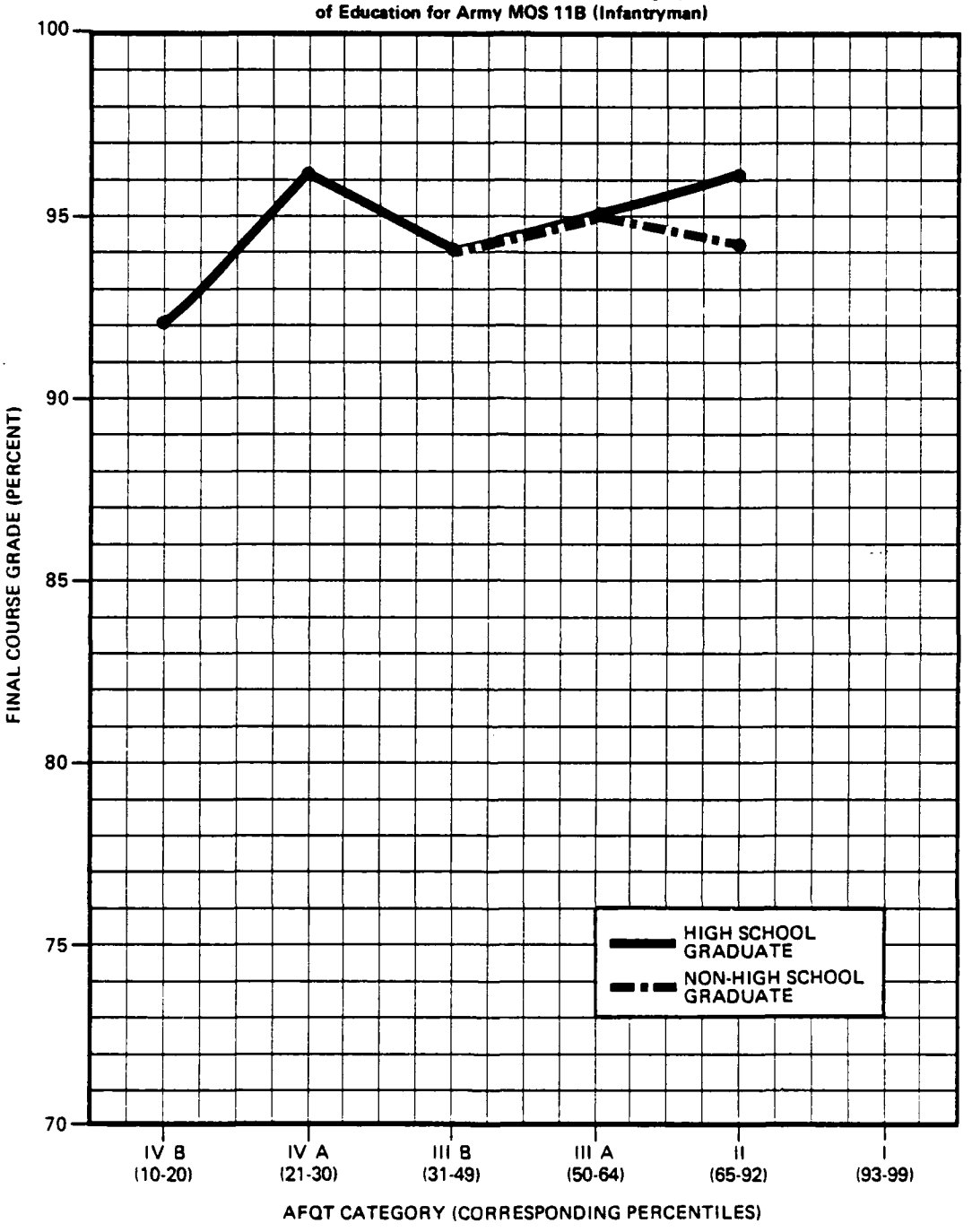


Figure 53
Final Course Grade as a Function of AFQT Category and Level of Education for Army MOS 11C (Indirect Fire Infantryman)

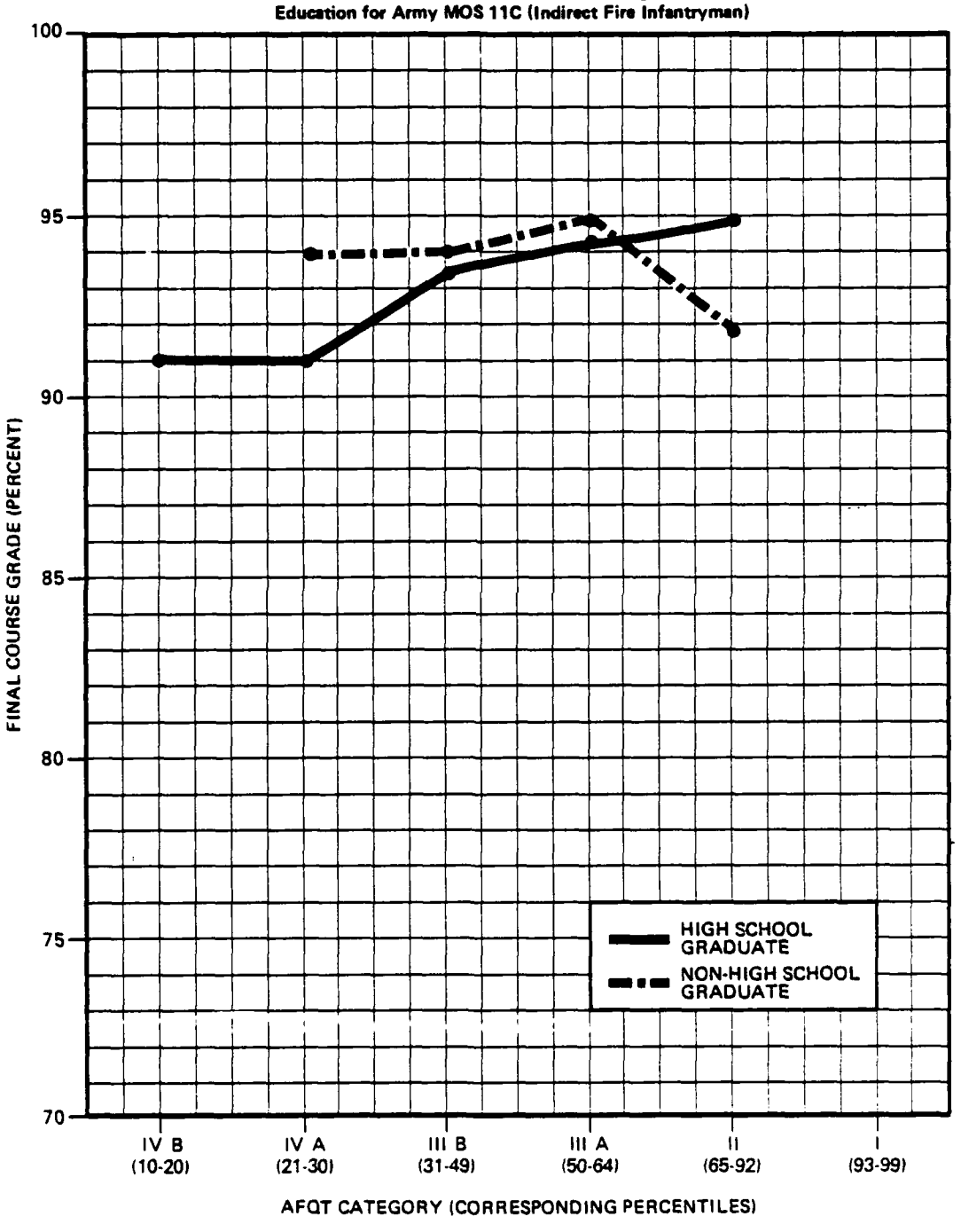


Figure 54
Final Course Grade as a Function of AFQT Category and Level of Education for Army MOS 19E (Armor Crewman)

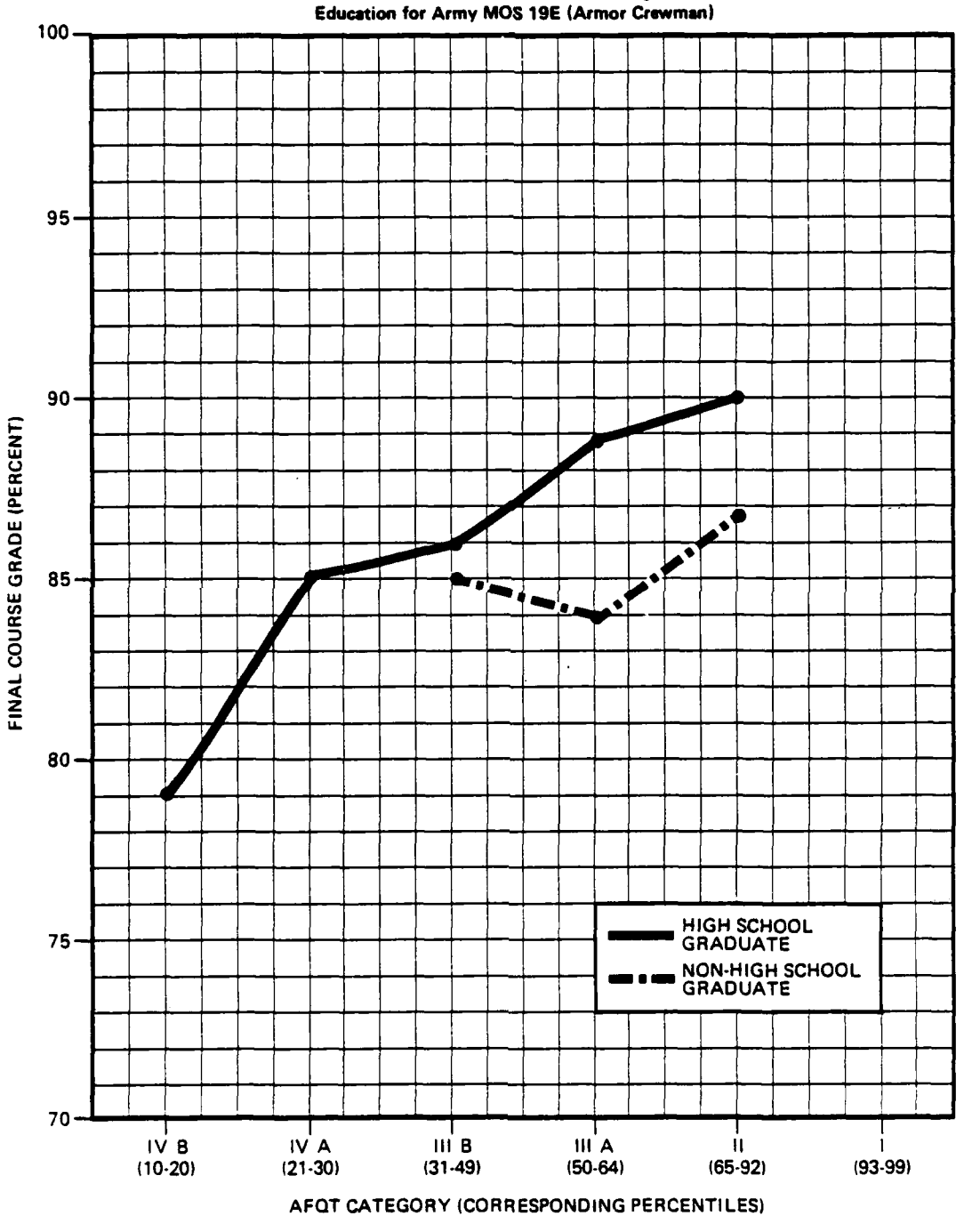


Figure 55
Final Course Grade as a Function of AFQT Category and Level of Education for Army MOS 05C (Radio Teletypewriter Operator)

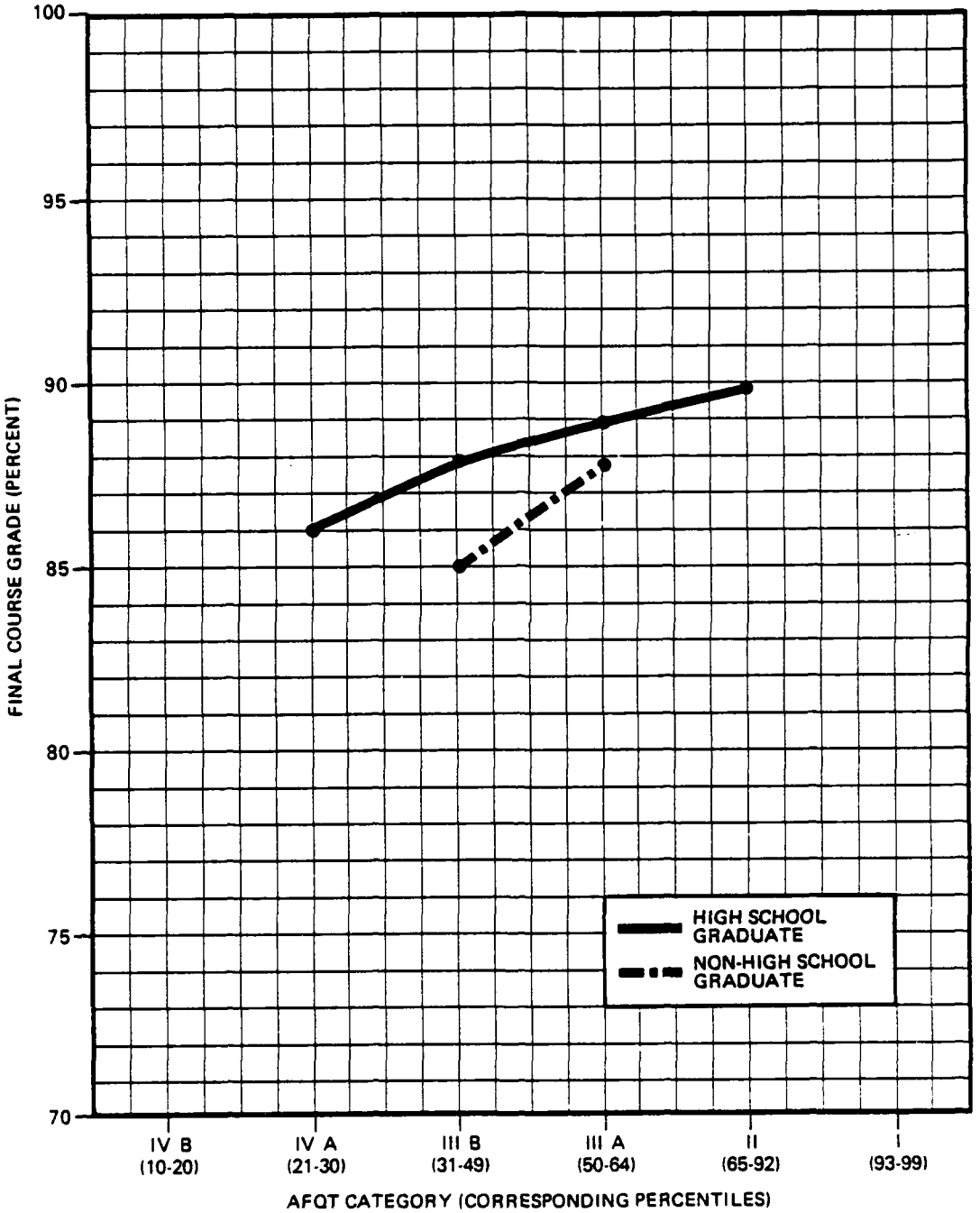


Figure 56
Final Course Grade as a Function of AFQT Category and Level of Education for Army MOS 31M (Multichannel Communications Operator)

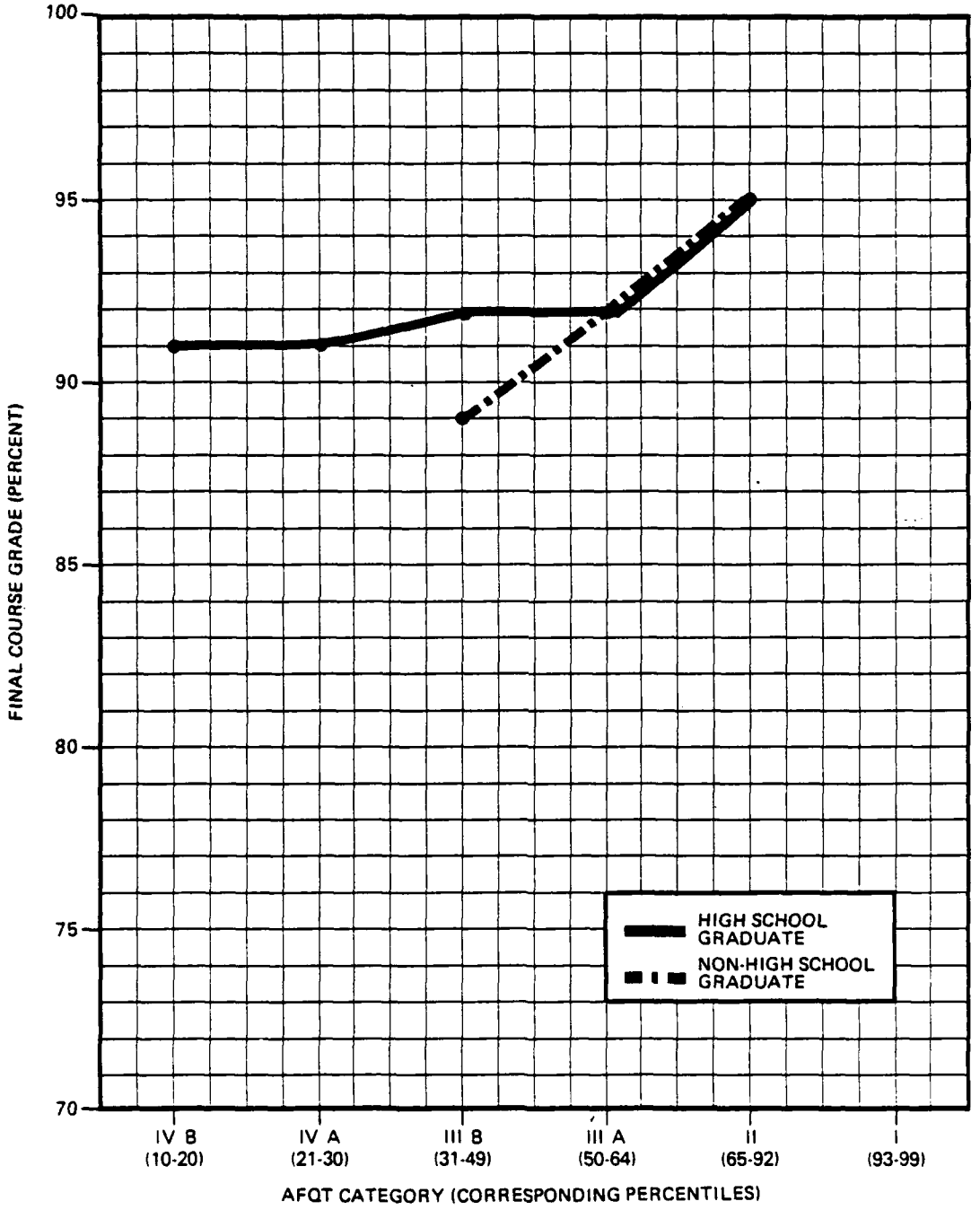


Figure 57
Final Course Grade as a Function of AFQT Category and Level of Education for Marine Corps Specialty 2841 (Ground Radio Repair) Basic Electronics Course

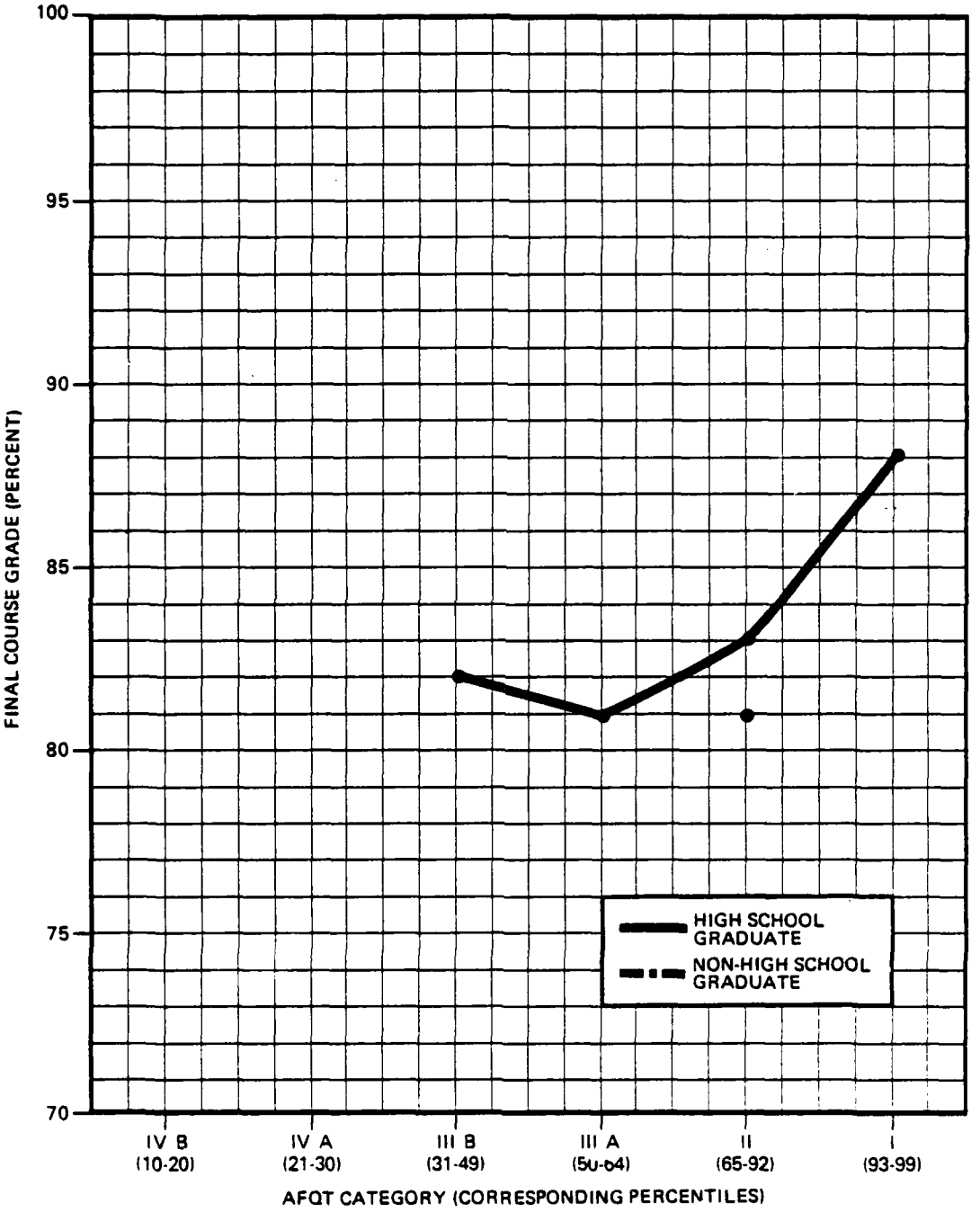


Figure 58
Final Course Grade as a Function of AFQT Category and Level of Education for Marine Corps Specialty 2841 (Ground Radio Repair) Radio Fundamentals Course

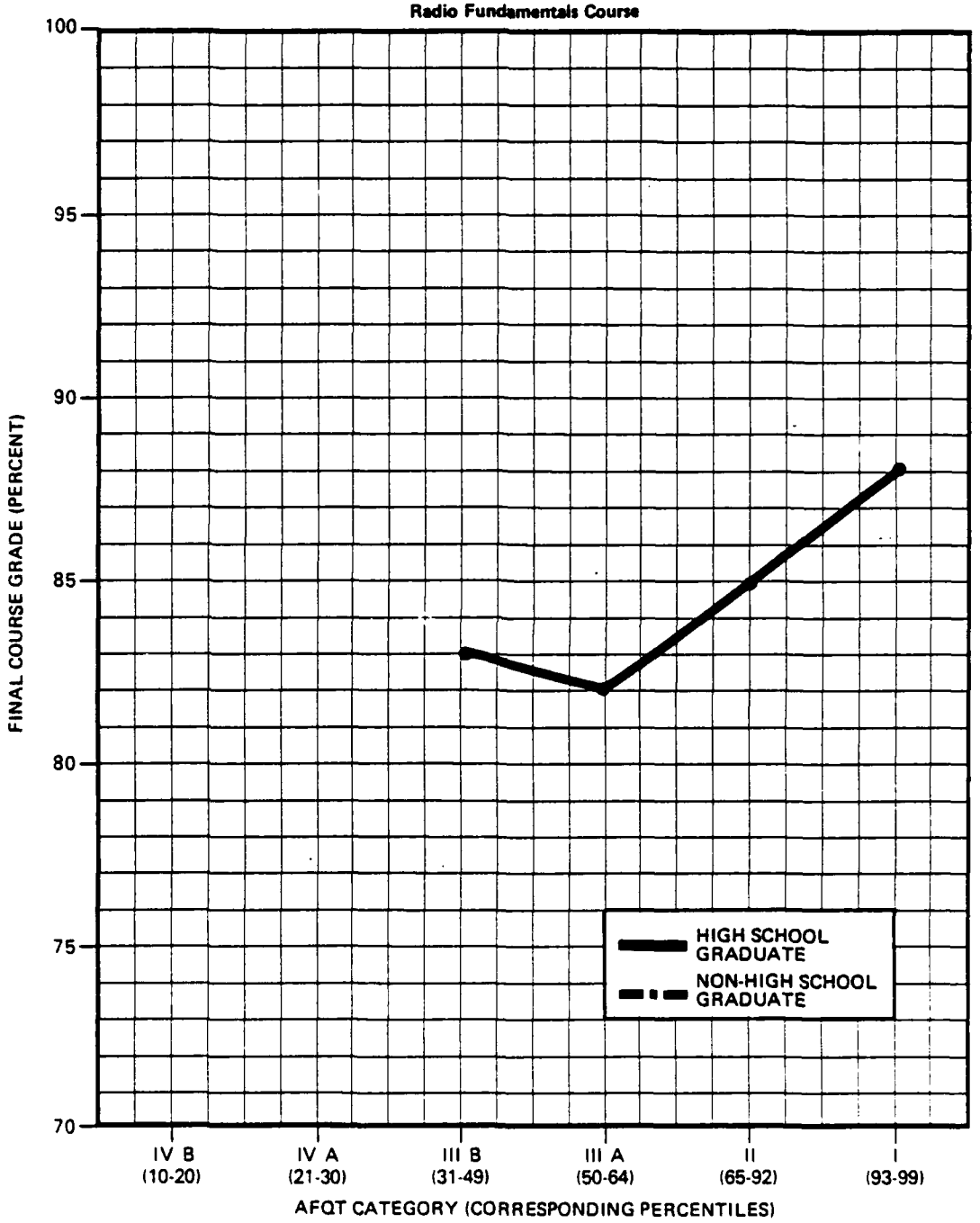


Figure 59
Final Course Grade as a Function of AFQT Category and Level of Education for Marine Corps Specialty 2841 (Ground Radio Repair) Ground Radio Repair Course

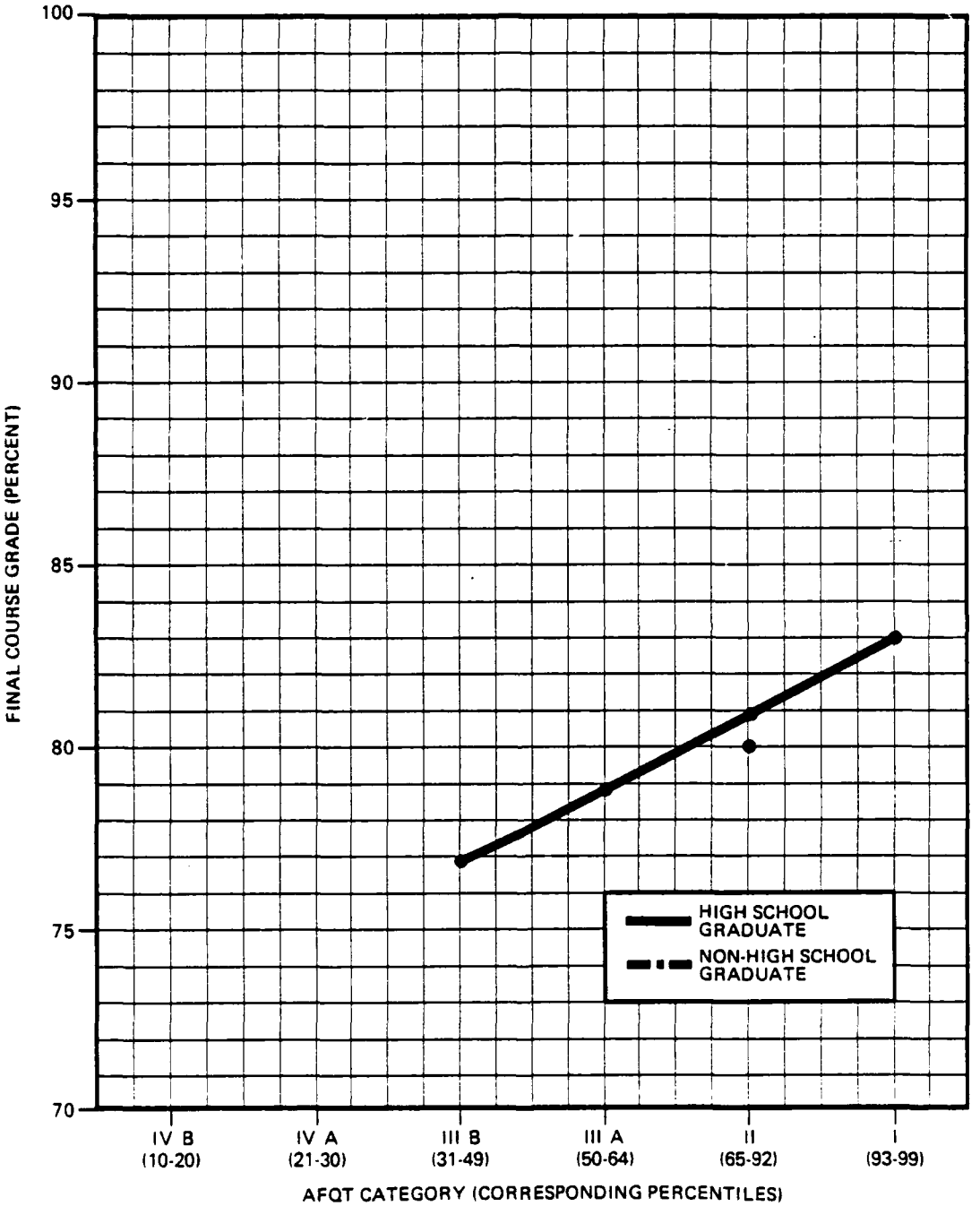


Figure 60
Final Course Grade as a Function of AFQT Category and Level of Education for Army MOS 67N (Utility Helicopter Repairer)

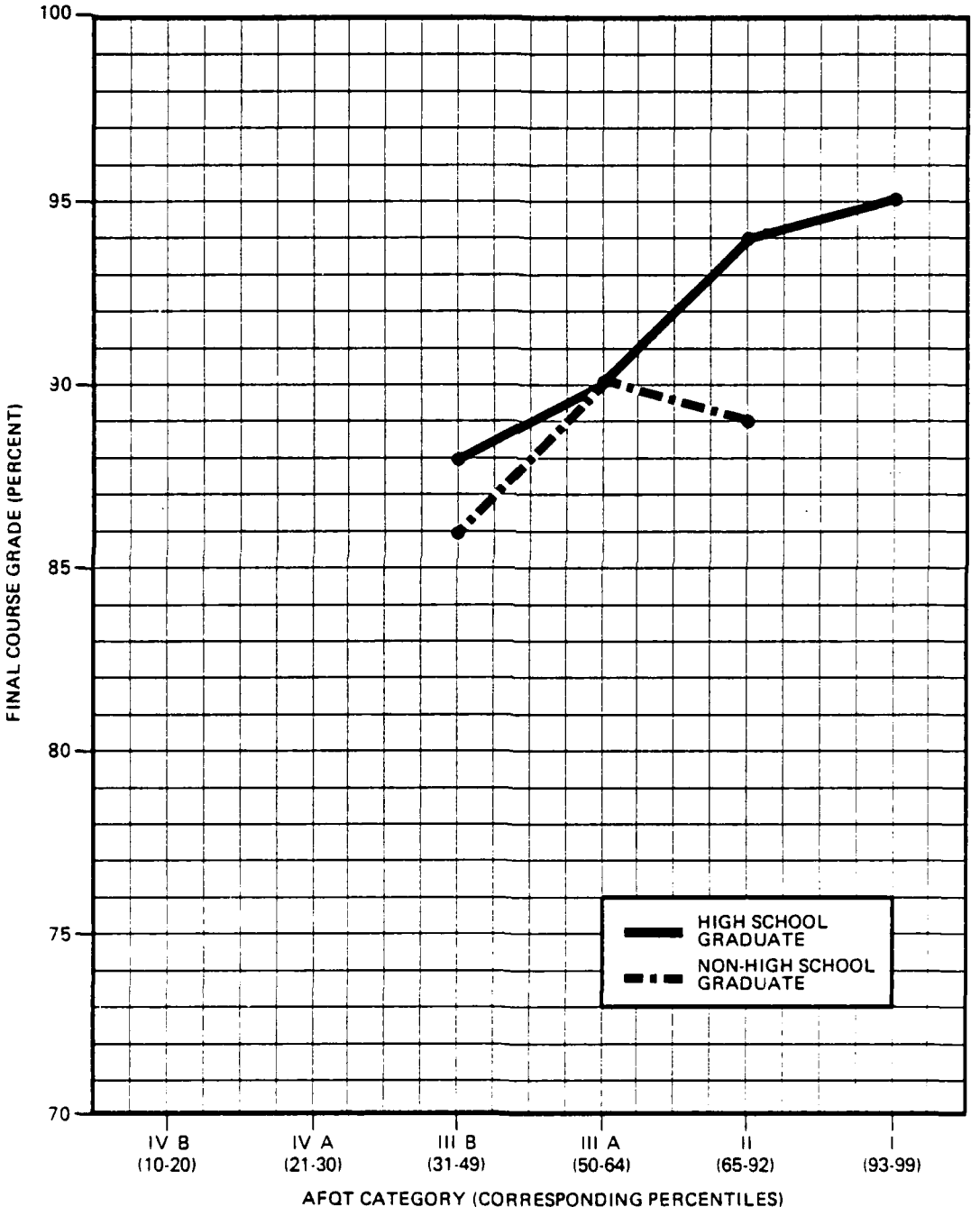


Figure 61
Final Course Grade as a Function of AFQT Category and Level of Education for Army MOS 73C (Finance Specialist)

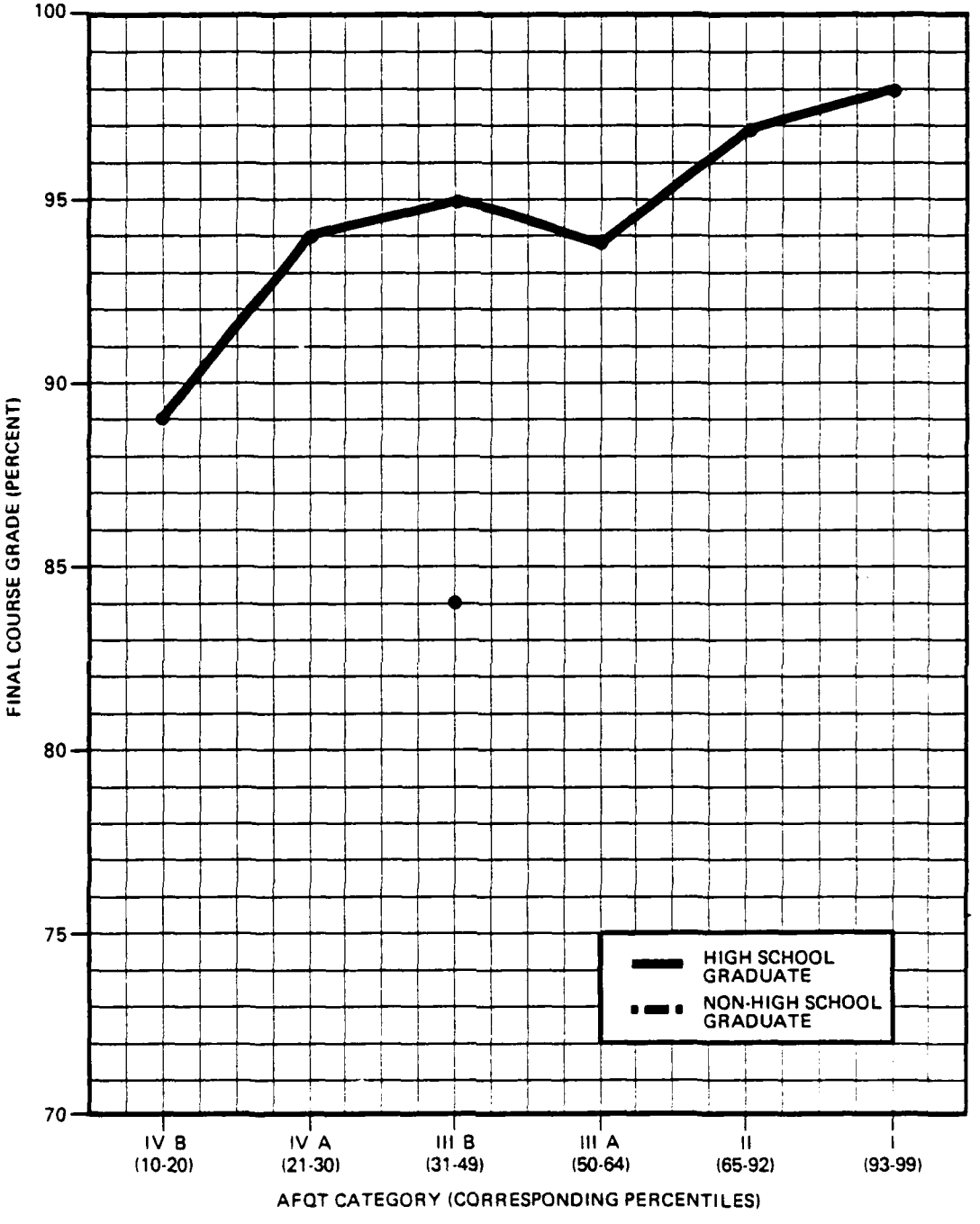


Figure 62
Final Course Grade as a Function of AFQT Category and Level of Education for Army MOS 75B (Personnel Administration Specialist)

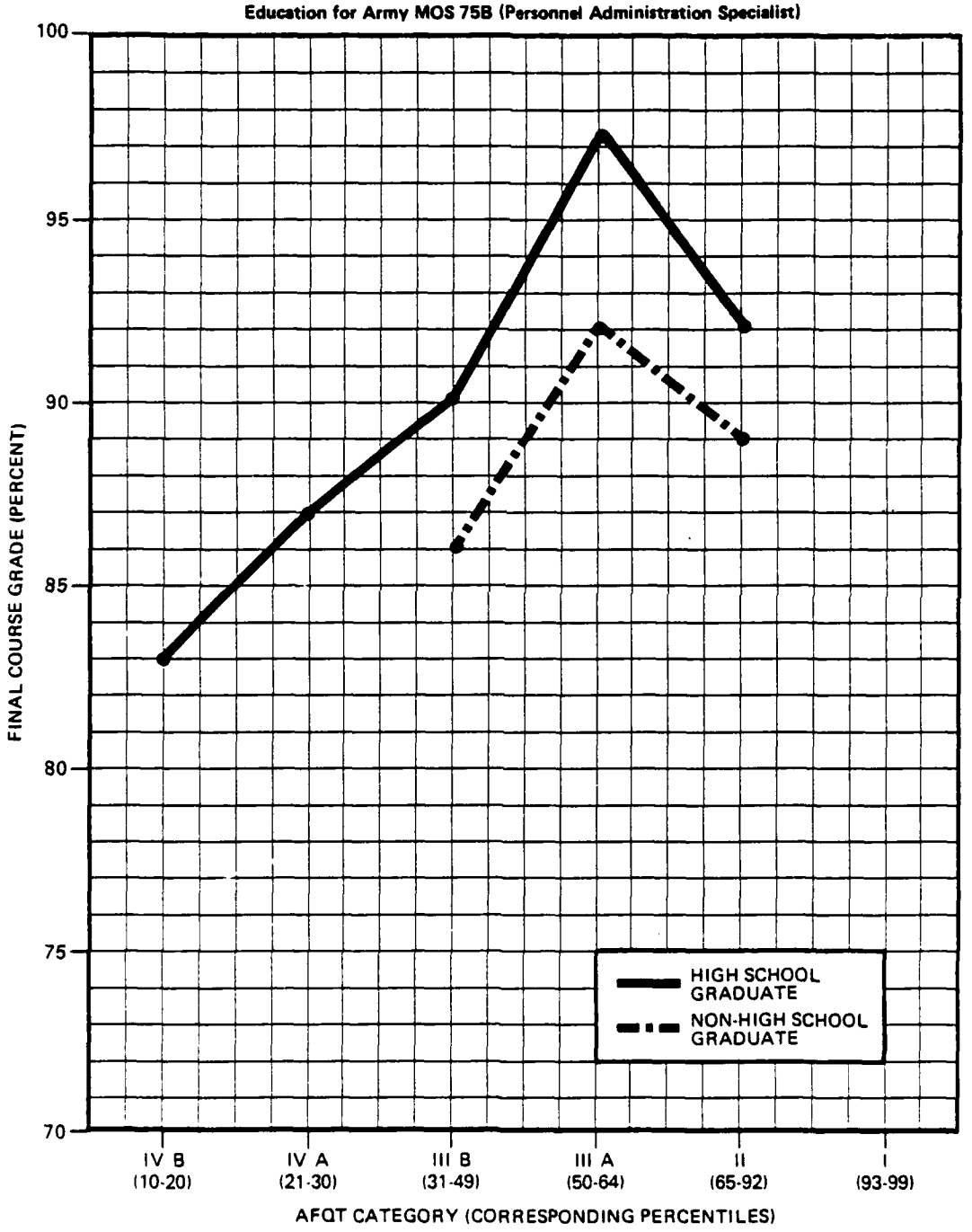


Figure 63
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Marine Corps Specialty 0311 (Infantryman)

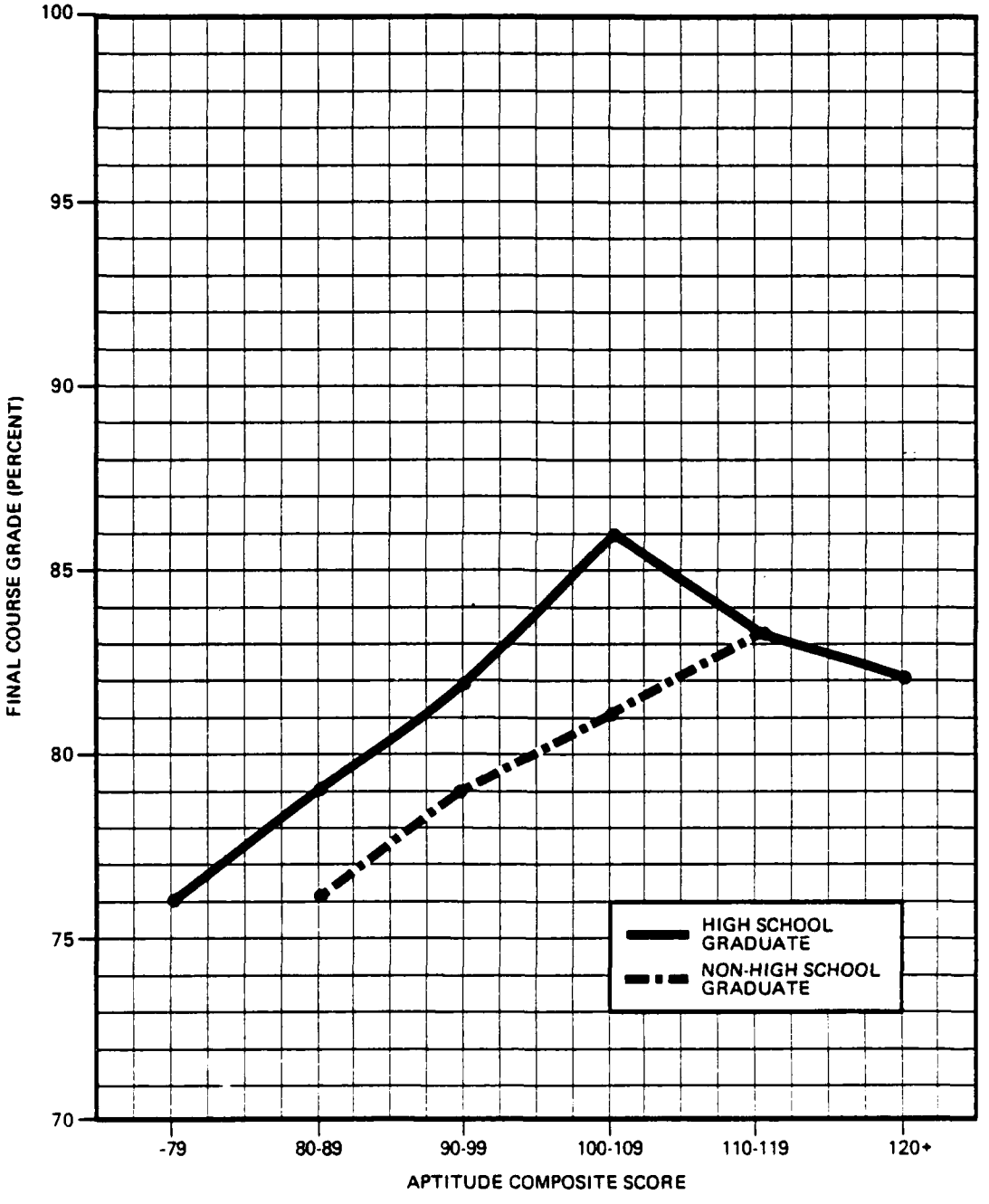


Figure 64
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Army MOS 11B (Infantryman)

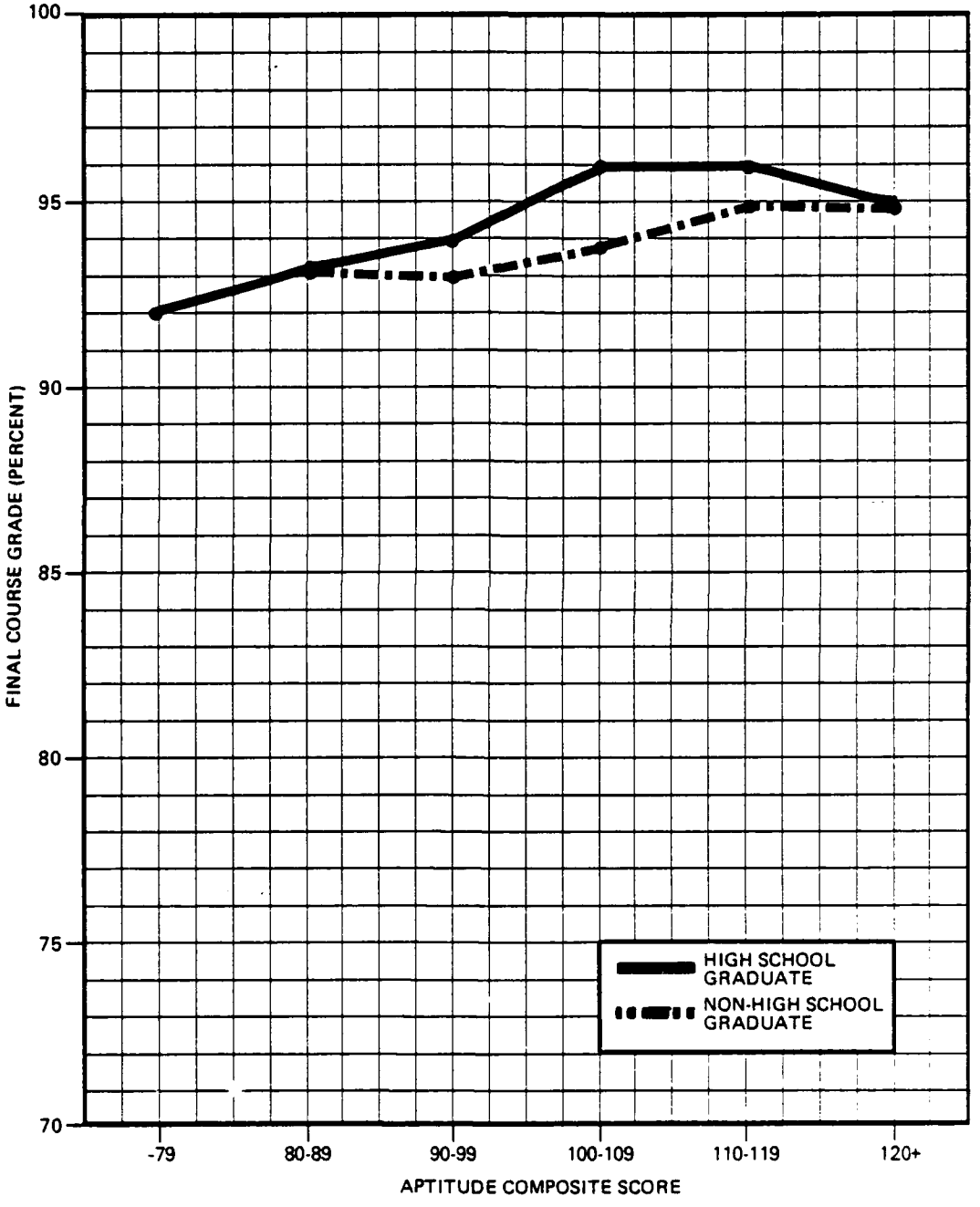


Figure 65
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Army MOS 11C (Indirect Fire Infantryman)

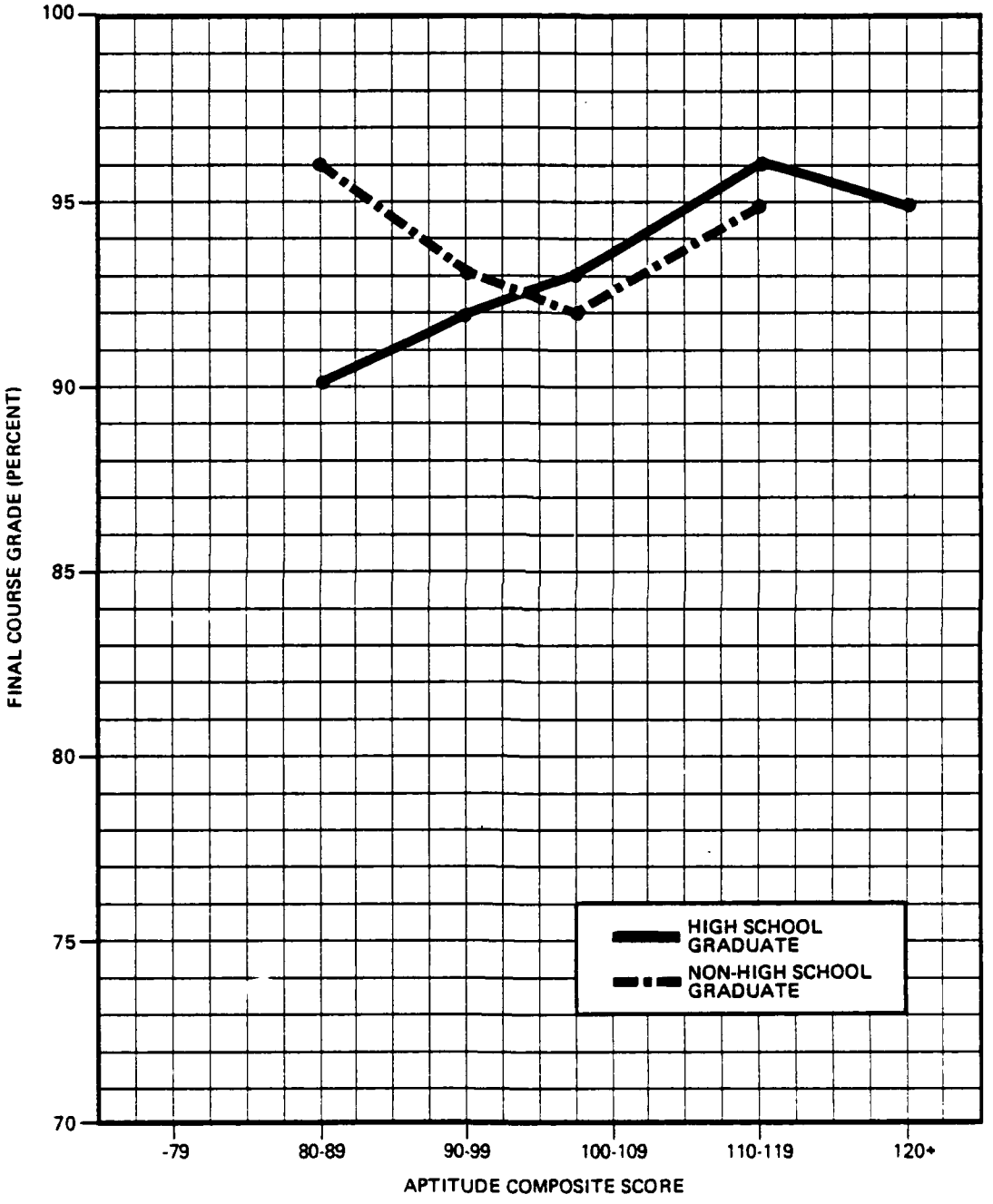


Figure 66
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Army MOS 19E (Armor Crewman)

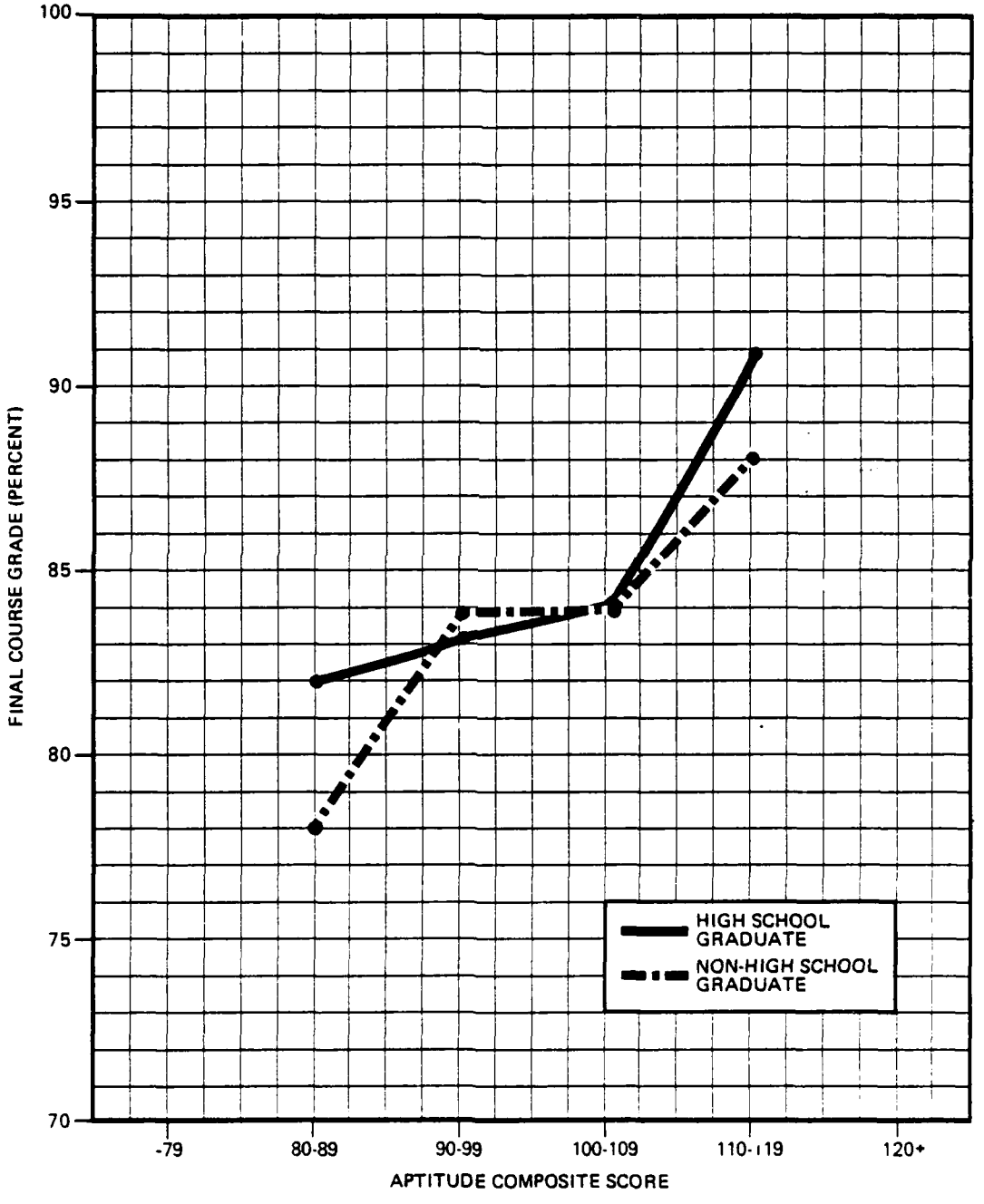


Figure 67
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Army MOS 05C (Radio Teletypewriter Operator)

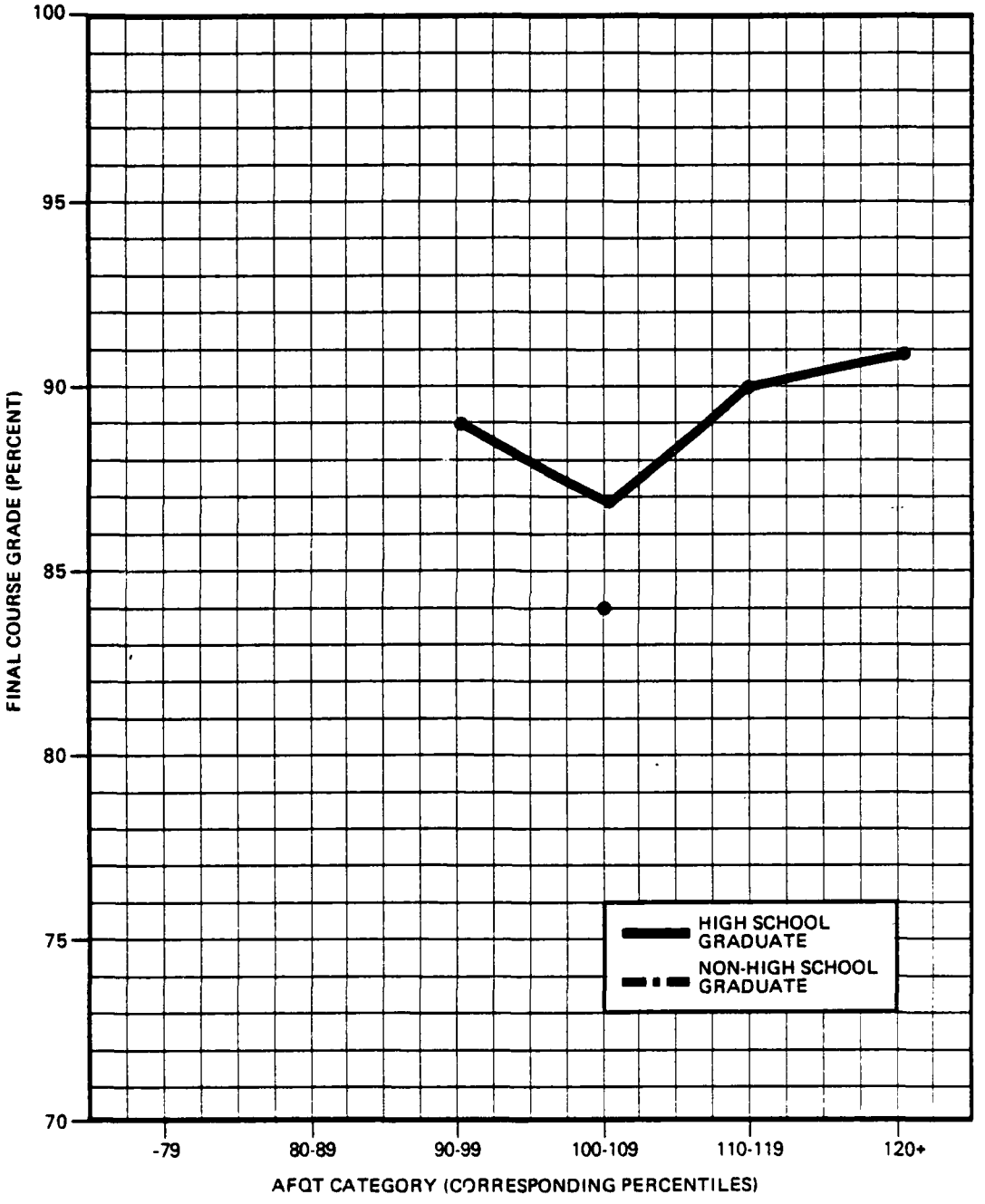


Figure 68
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Army MOS 31M (Multichannel Communications Operator)

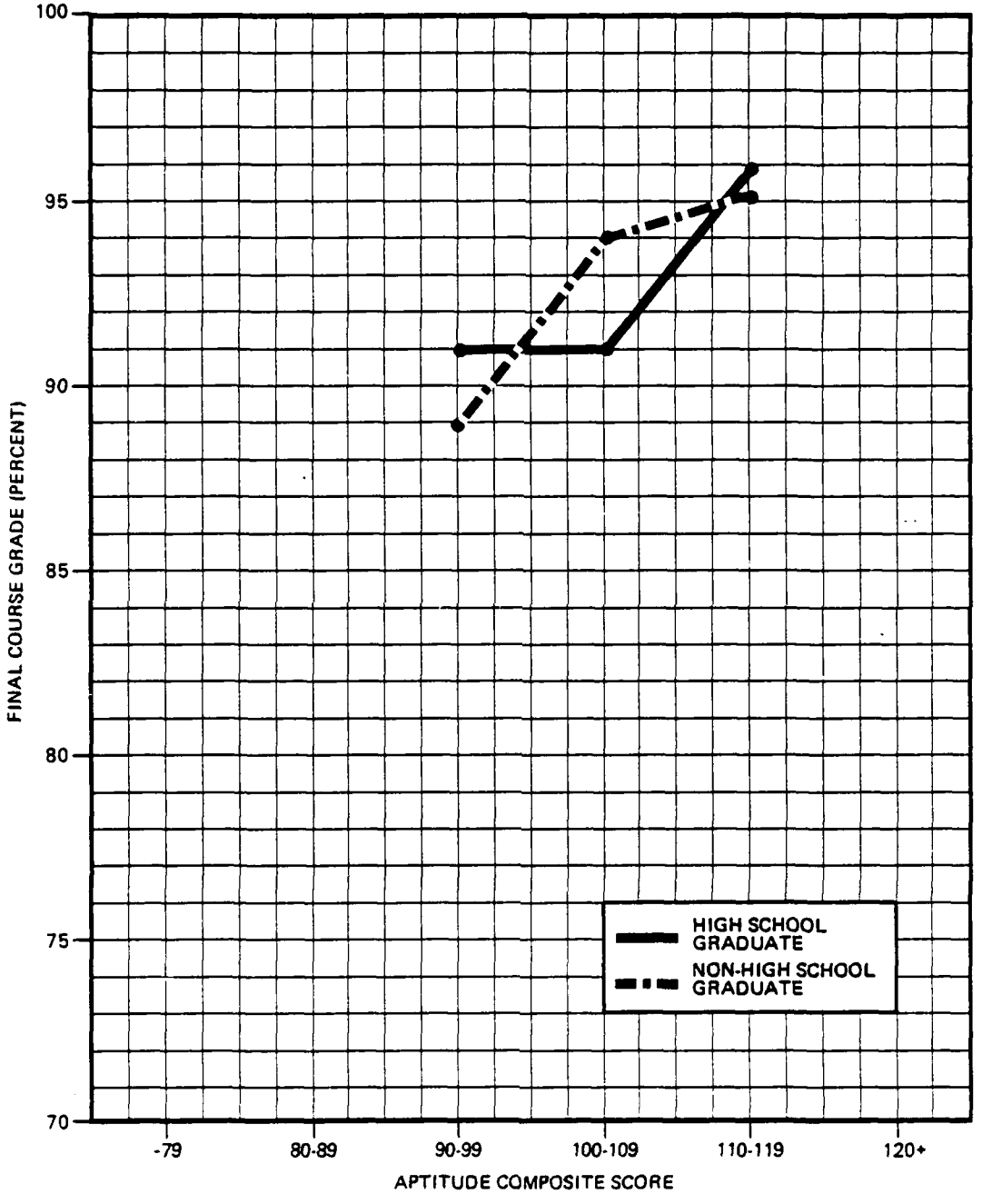


Figure 69
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Marine Corps Specialty 2841 (Ground Radio Repair) Basic Electronics Course

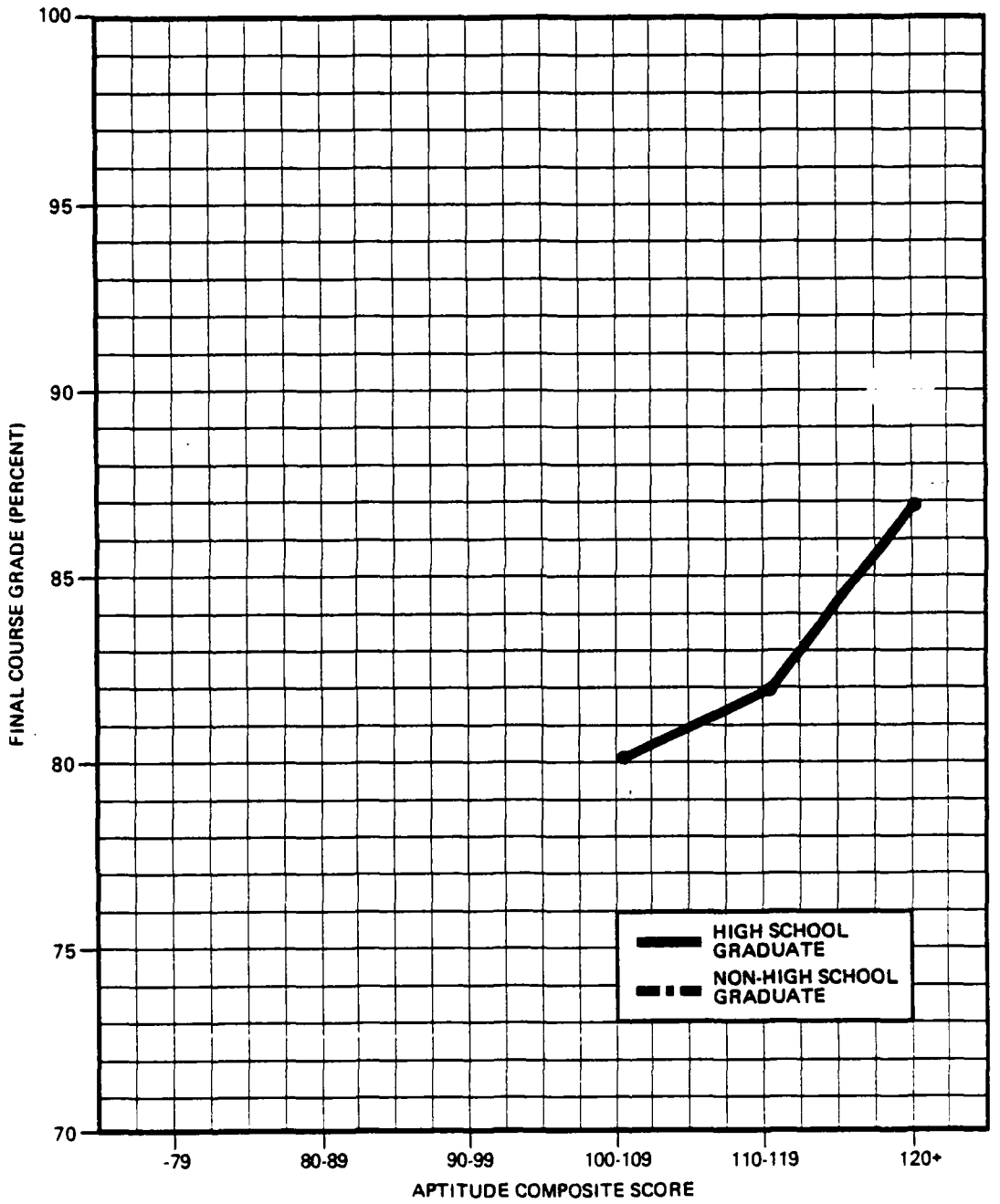


Figure 70
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Marine Corps Specialty 2841 (Ground Radio Repair)
Radio Fundamentals Course

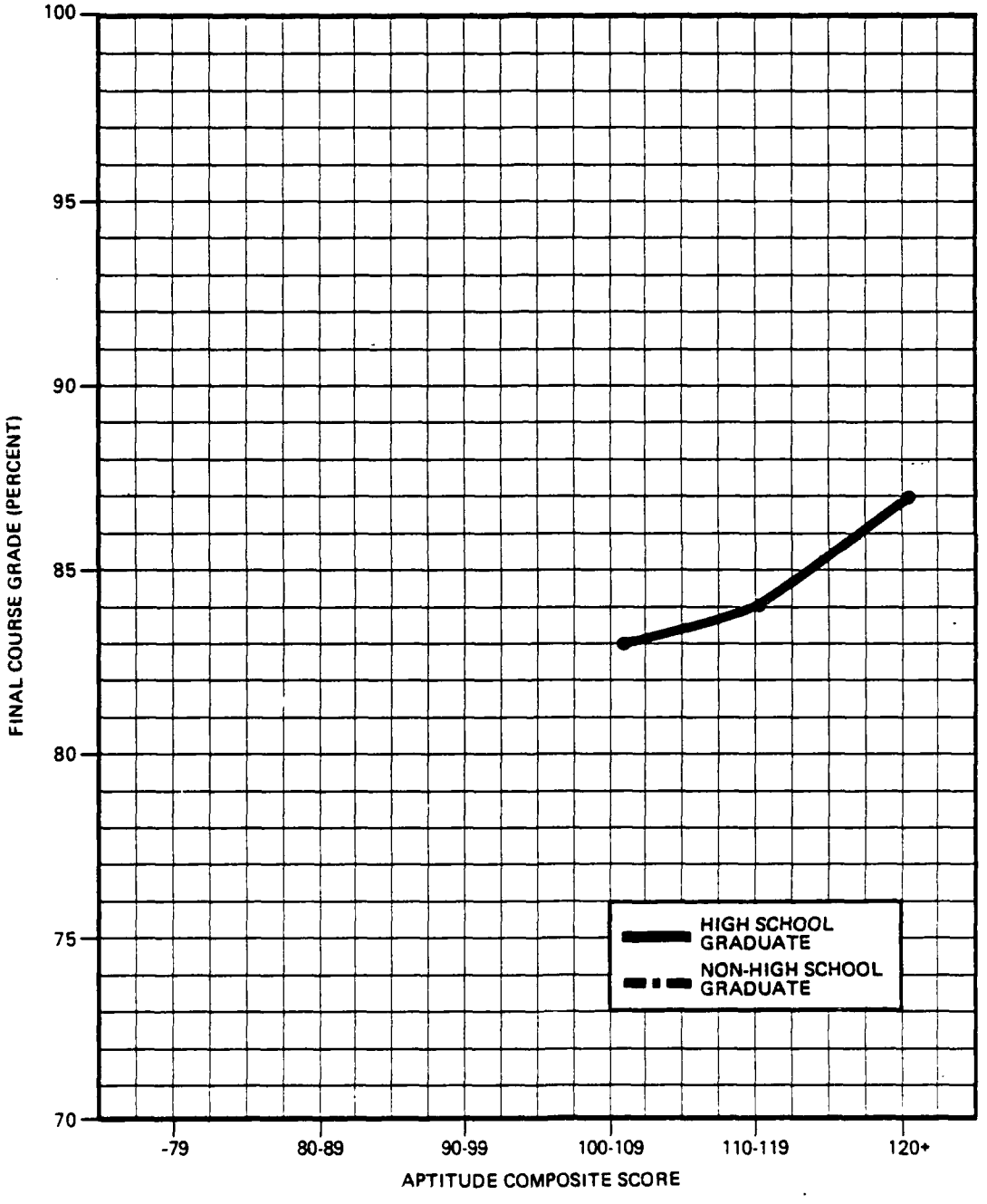


Figure 71
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Marine Corps Specialty 2841 (Ground Radio Repair)
Ground Radio Repair Course

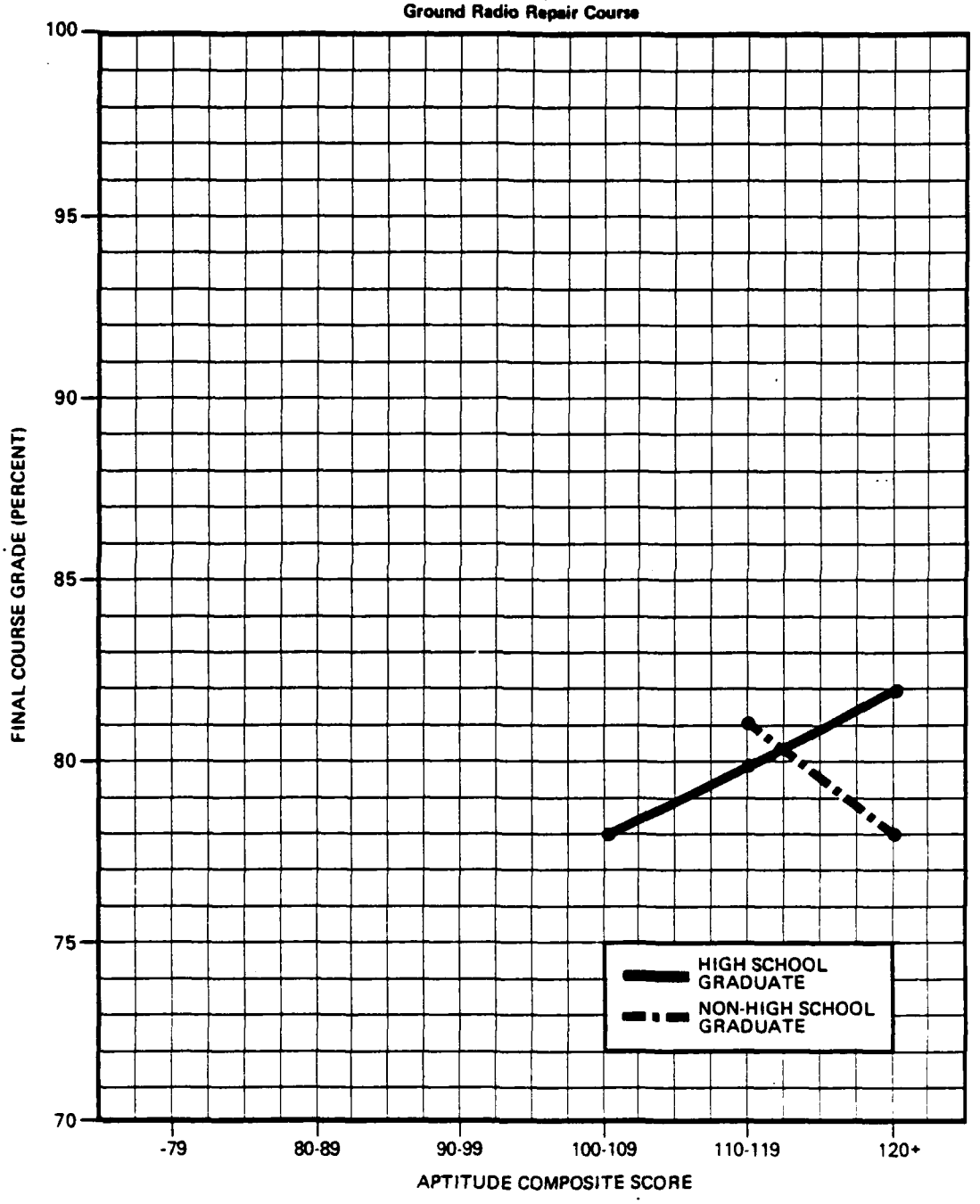


Figure 72
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Army MOS 67N (Utility Helicopter Repairer)

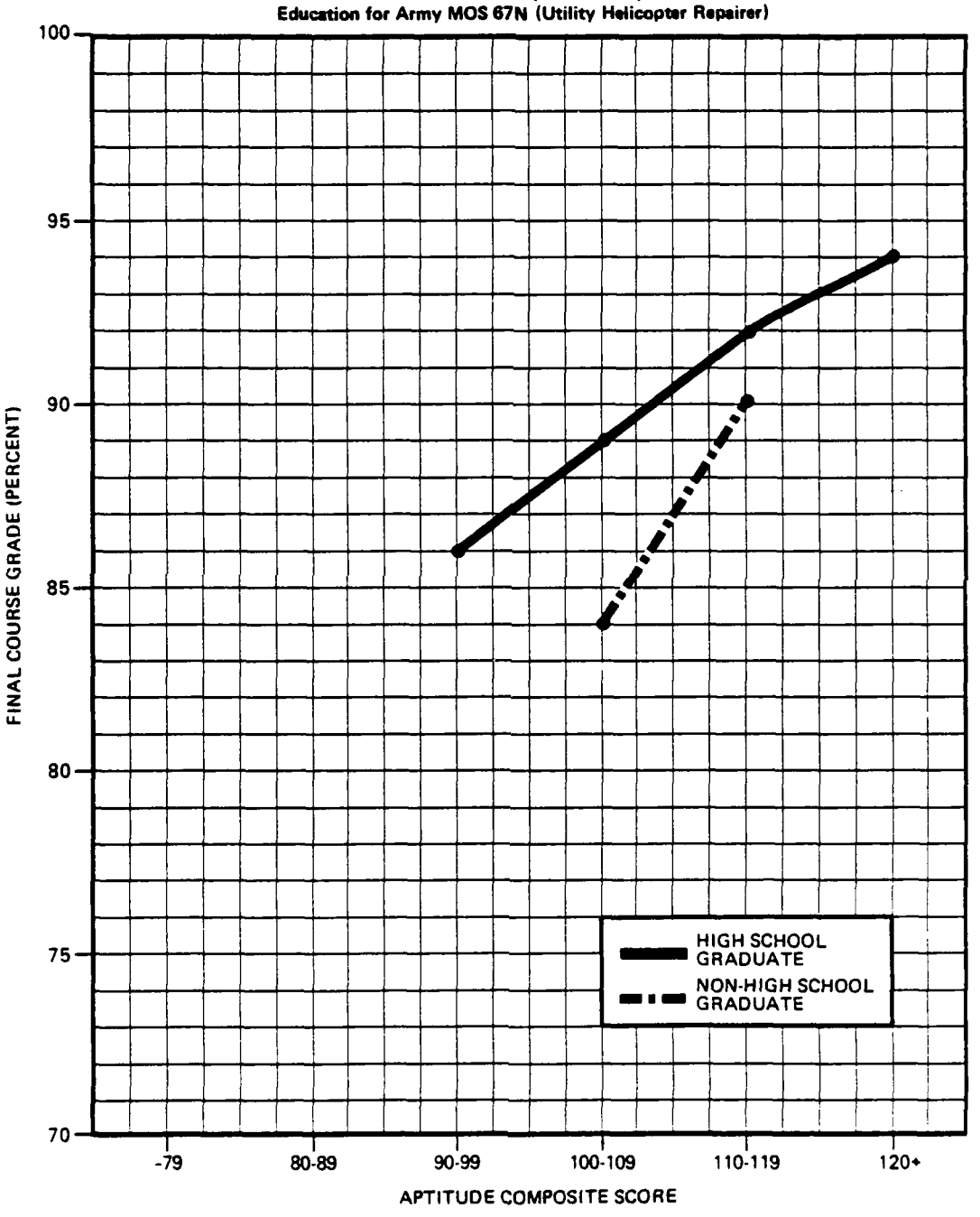


Figure 73
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Army MOS 73C (Finance Specialist)

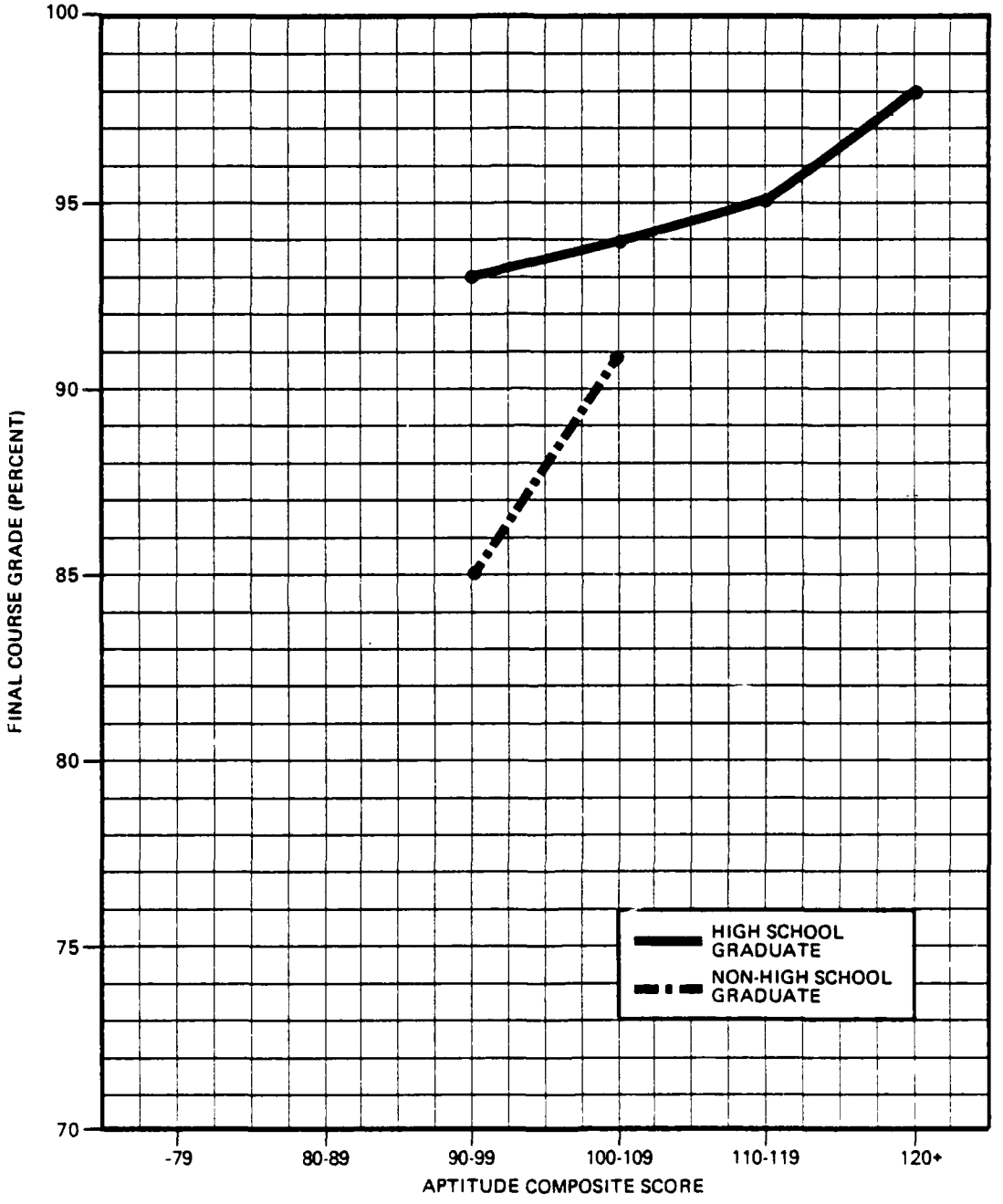


Figure 74
Final Course Grade as a Function of Aptitude Composite Score and Level of Education for Army MOS 75B (Personnel Administration Specialist)

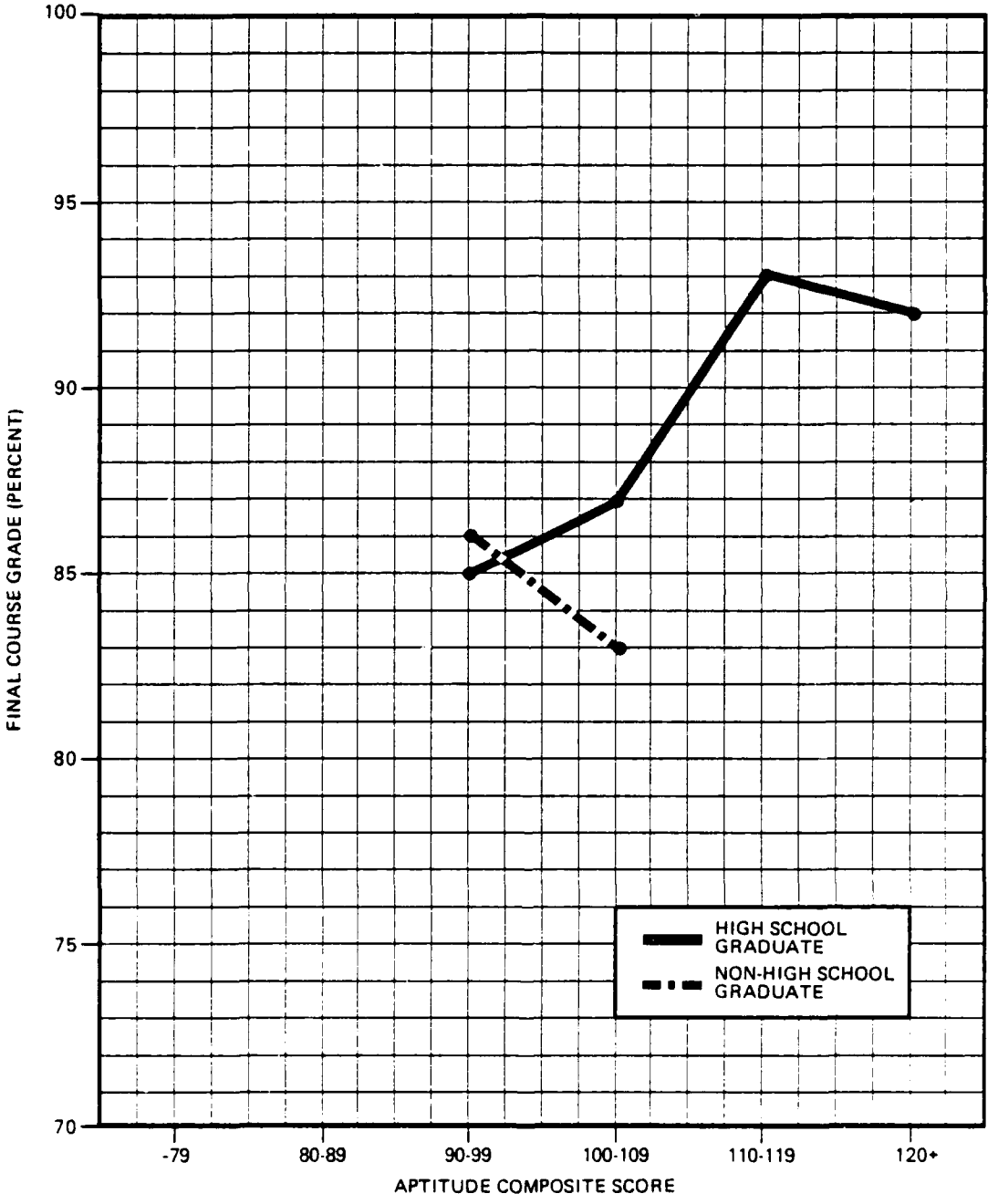


Figure 75
Final Course Grade as a Function of AFQT Score and Racial/Ethnic Group for Army MOS 05C (Radio Teletypewriter Operator)

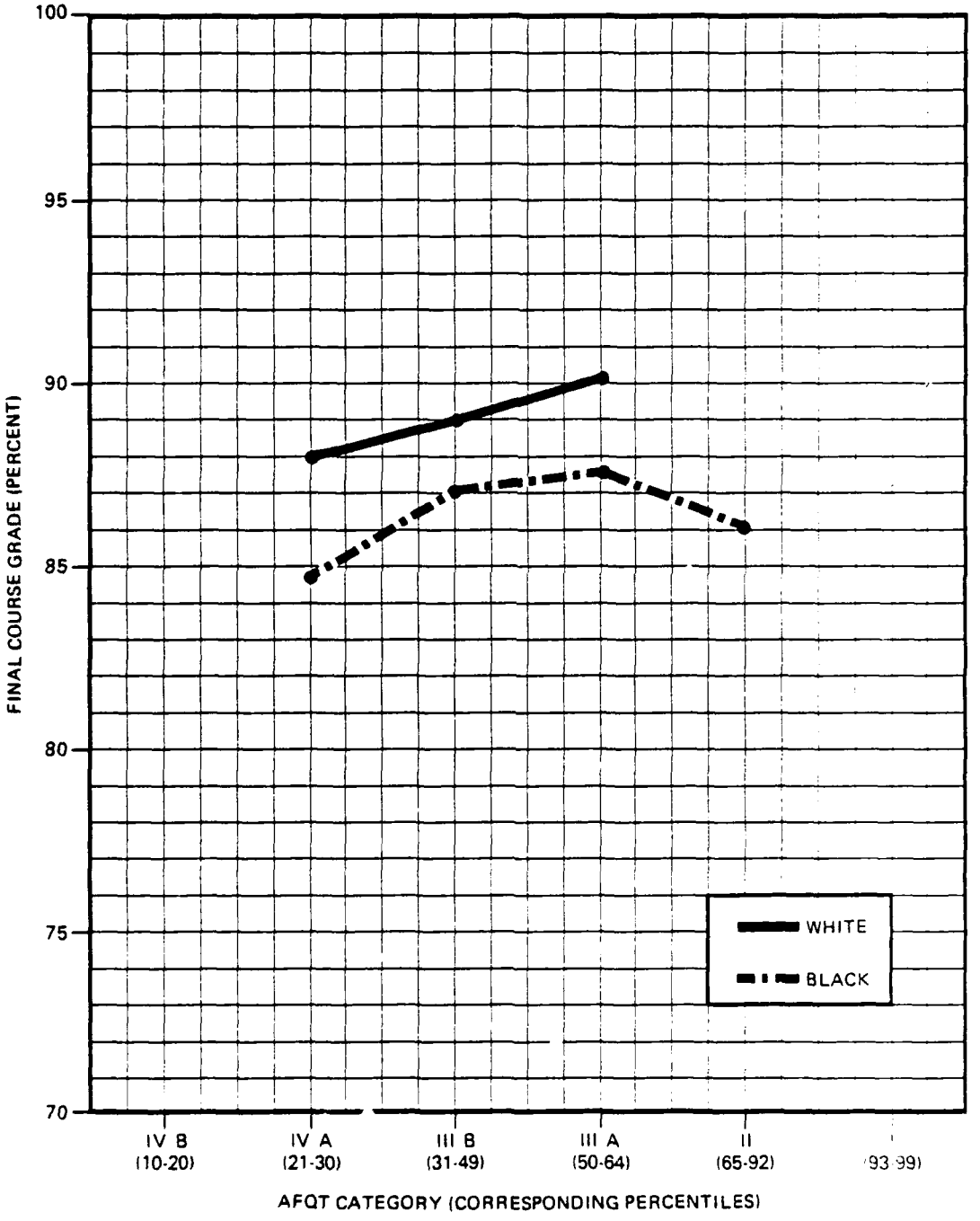


Figure 76
Final Course Grade as a Function of AFQT Score and Racial/Ethnic Group
for Army MOS 31M (Multichannel Communications Operator)

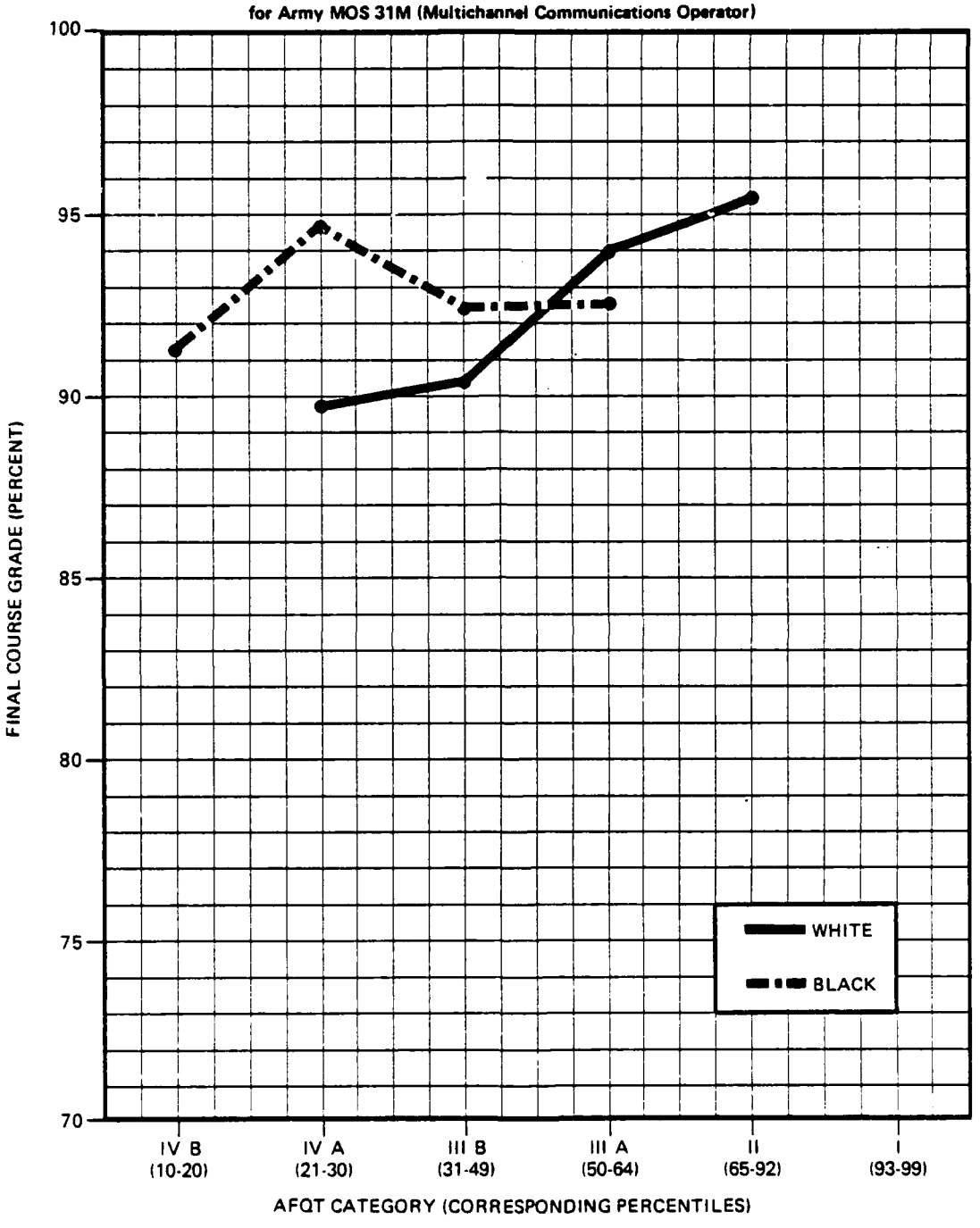


Figure 77
Final Course Grade as a Funciton of AFQT Score and Racial/Ethnic Group for Army MOS 73C (Finance Specialist)

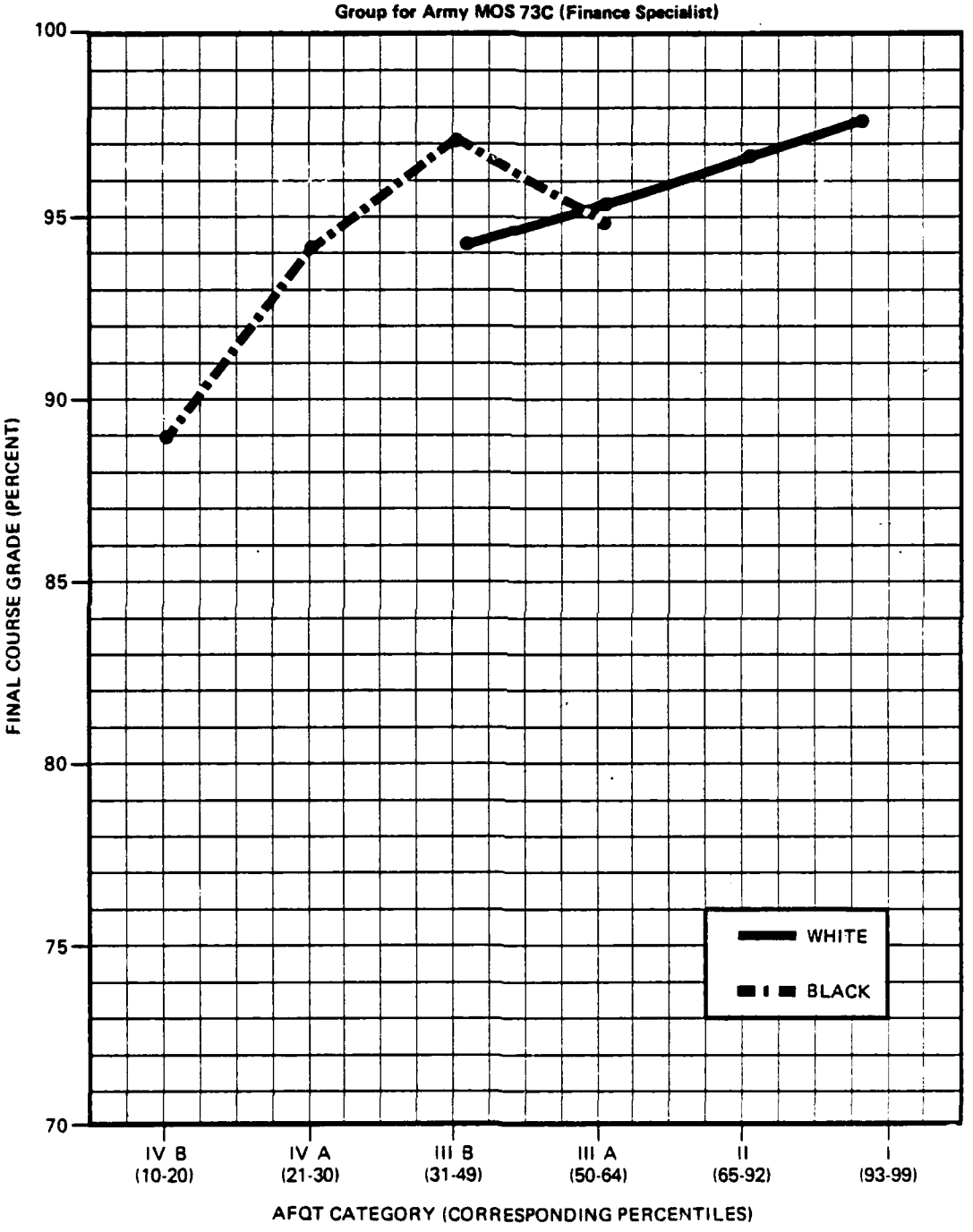


Figure 78
Final Course Grade as a Function of AFQT Score and Racial/Ethnic Group for Army MOS 75B (Personnel Administration Specialist)

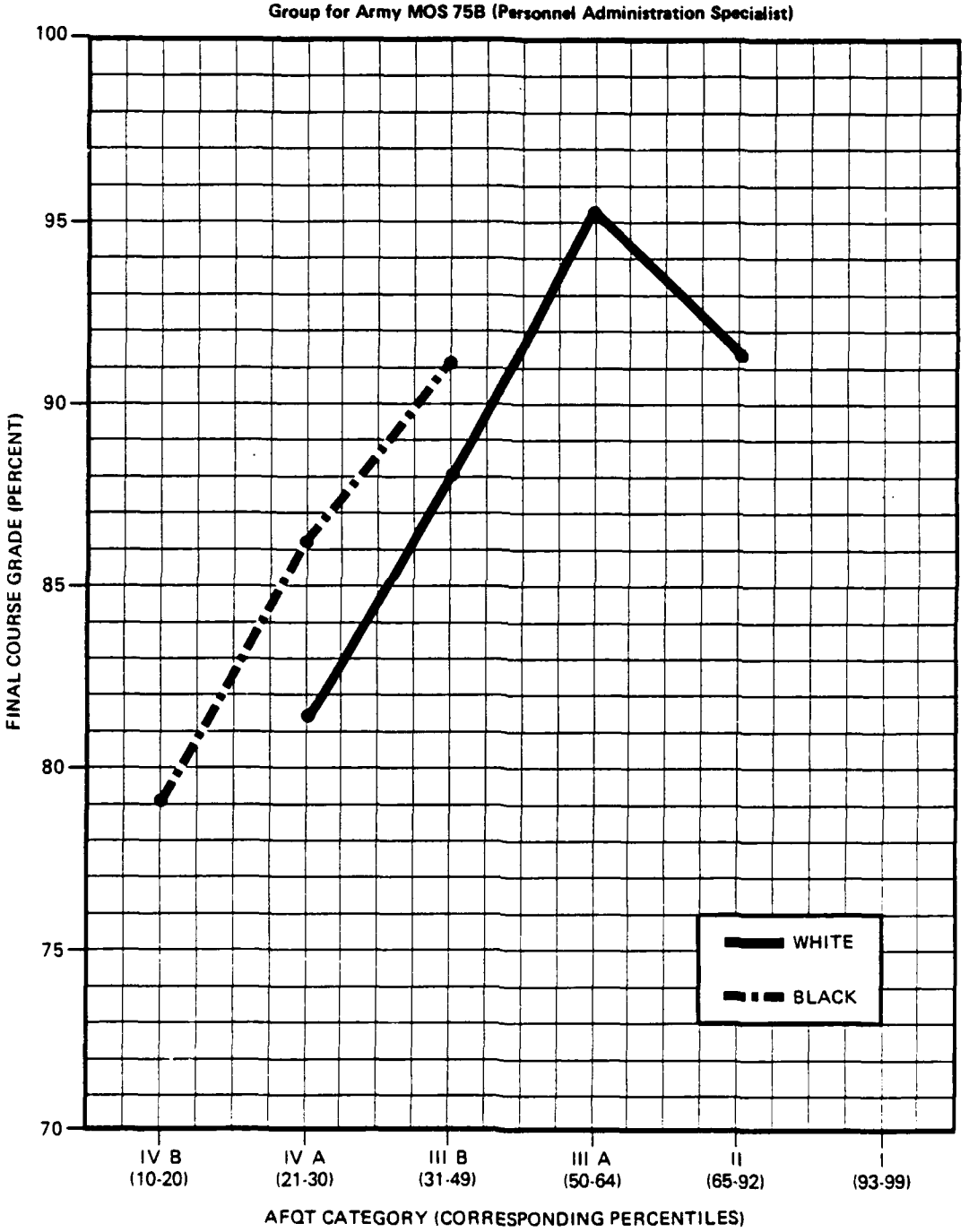


Figure 79
Final Course Grade as a Function of Aptitude Composite Score and Racial/Ethnic Group for Army MOS 05C (Radio Teletypewriter Operator)

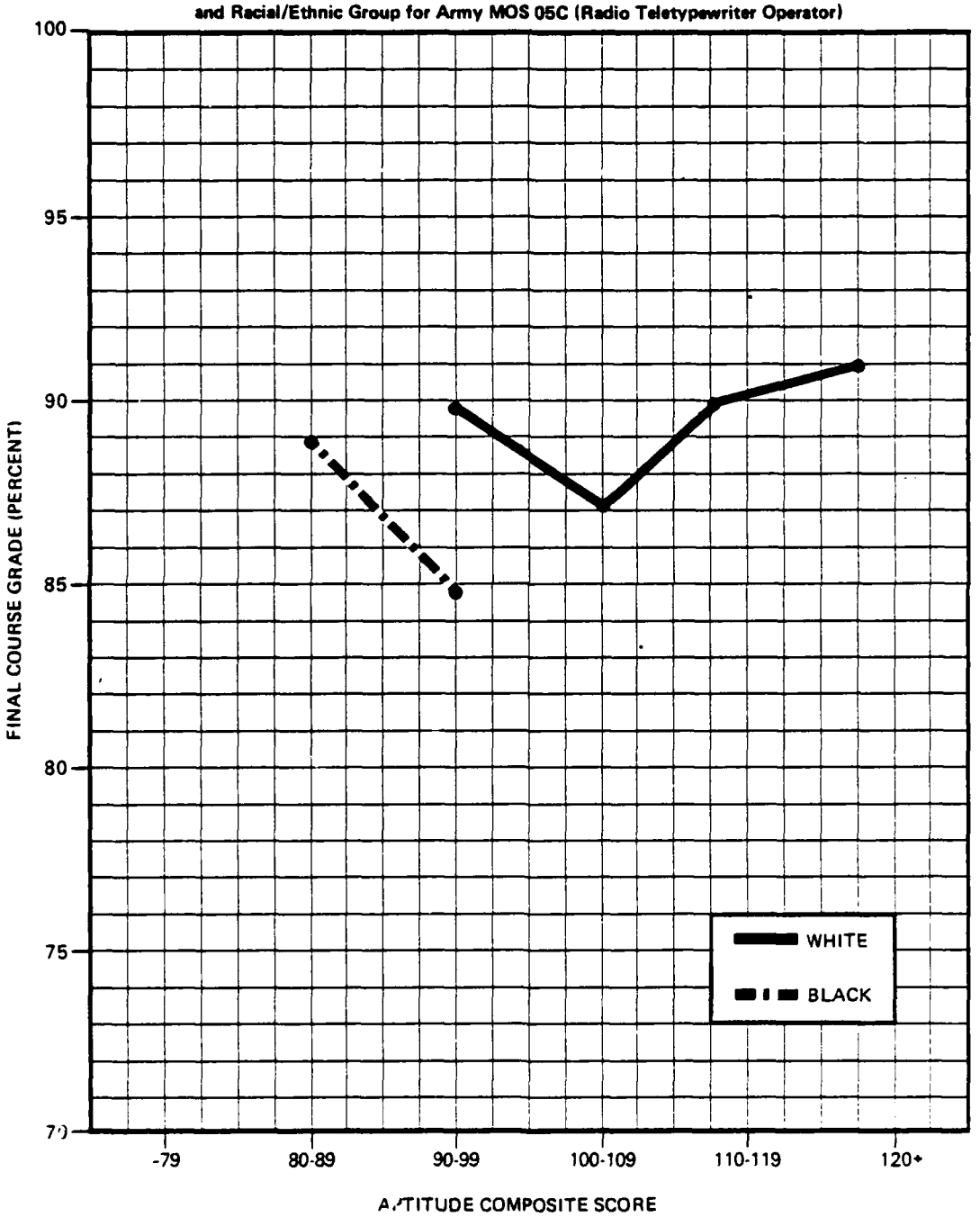


Figure 80
Final Course Grade as a Function of Aptitude Composite Score and Racial/Ethnic Group for Army MOS 31M (Multichannel Communications Operator)

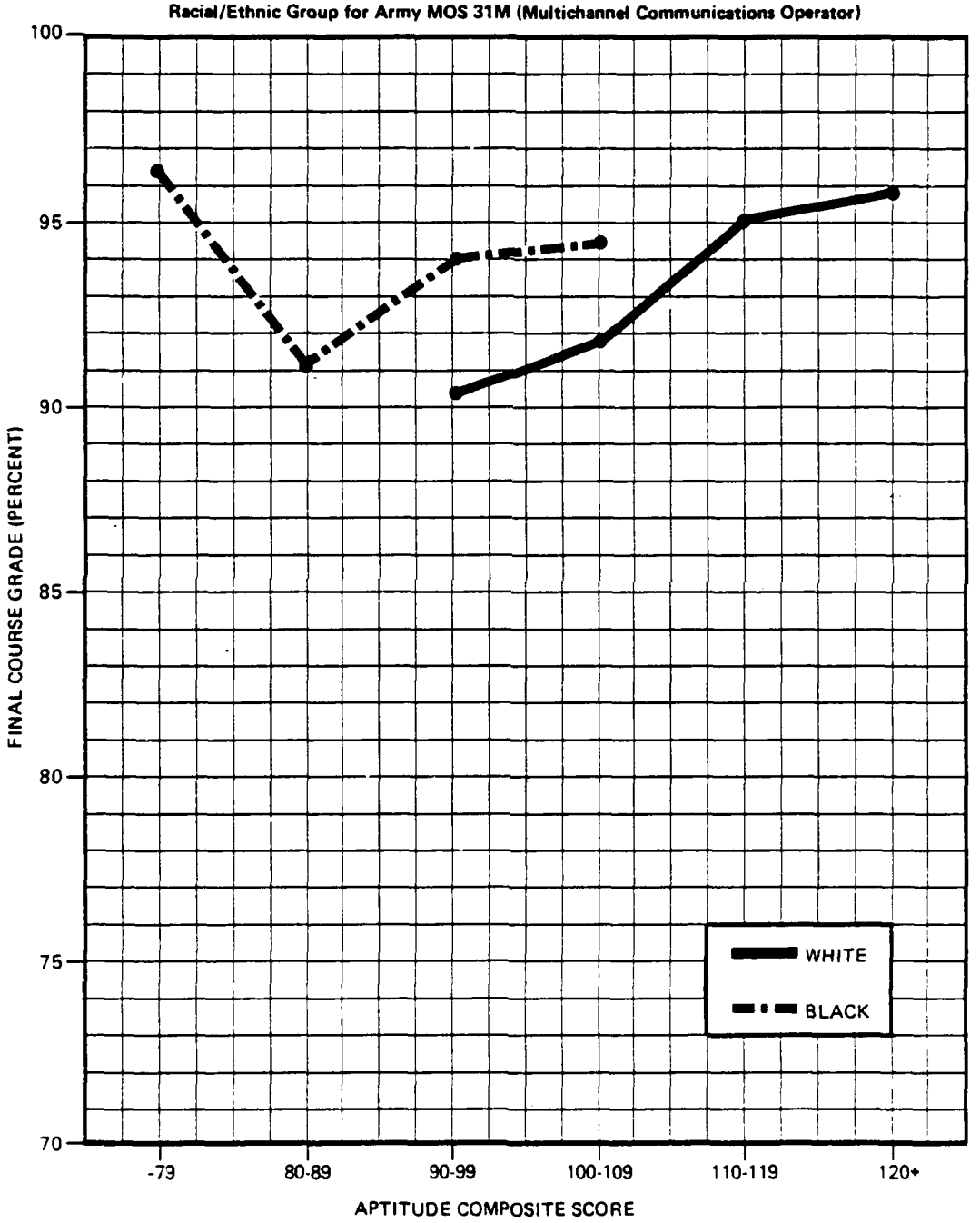


Figure 81
Final Course Grade as a Function of Aptitude Composite Score and Racial/Ethnic Group for Army MOS 73C (Finance Specialist)

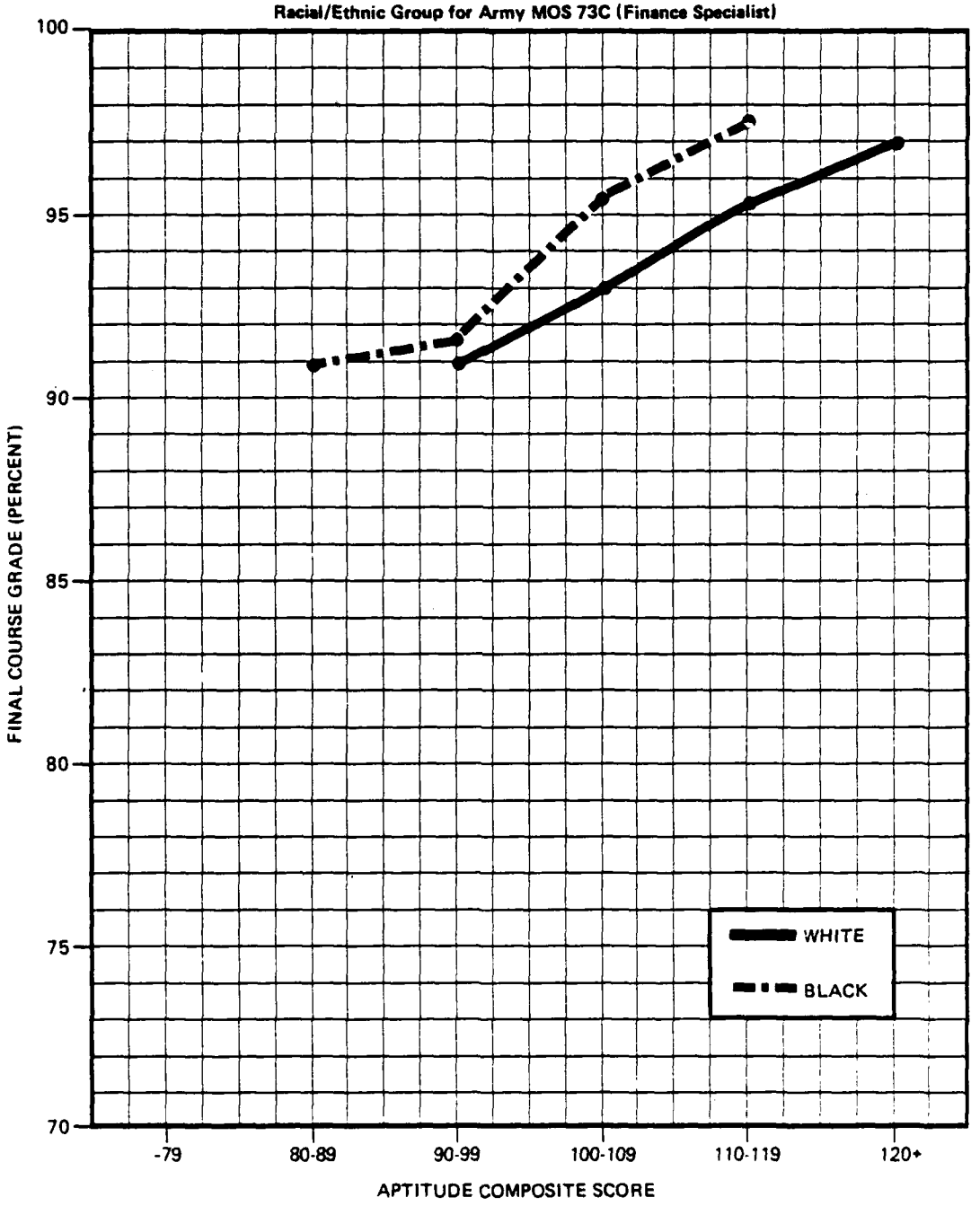


Figure 82
Final Course Grade as a Function of Aptitude Composite Score and Racial/Ethnic Group for Army MOS 75B (Personnel Administration Specialist)

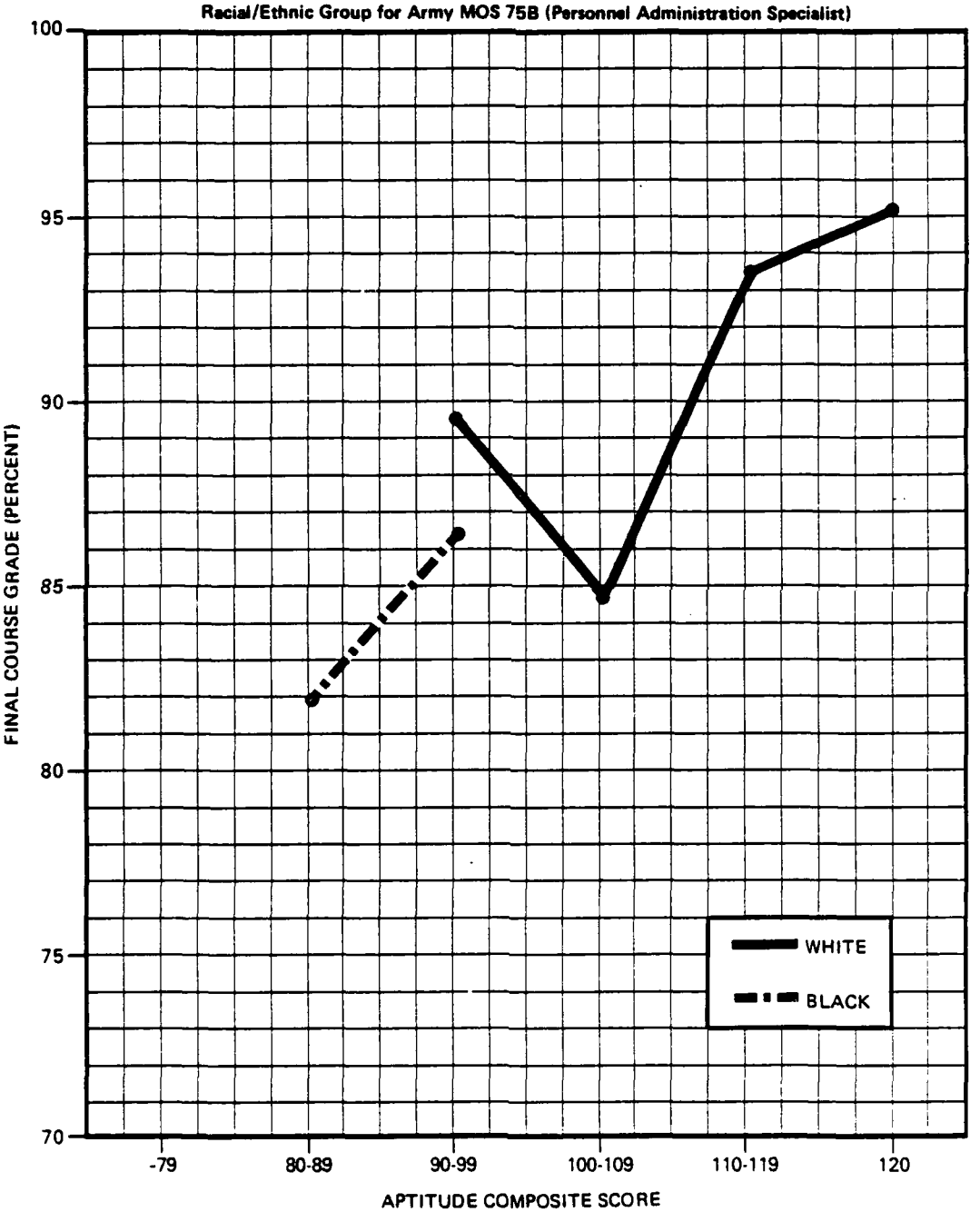


Figure 83
Percent of Attrition by AFQT Category for One Marine and Three Army Specialties

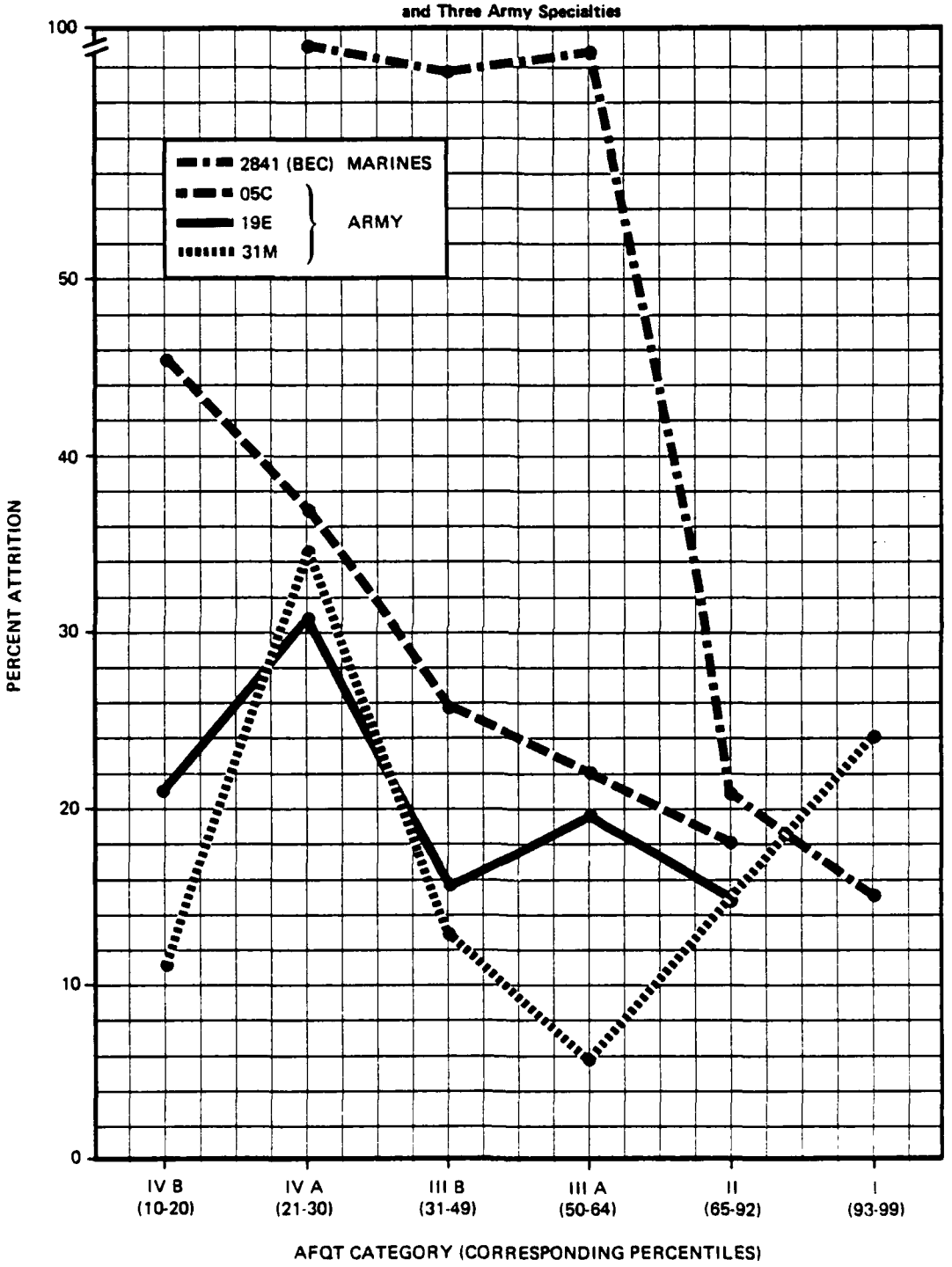


Figure 84
Percent of Attrition by Aptitude Composite Score for
One Marine Corps and Three Army Specialties

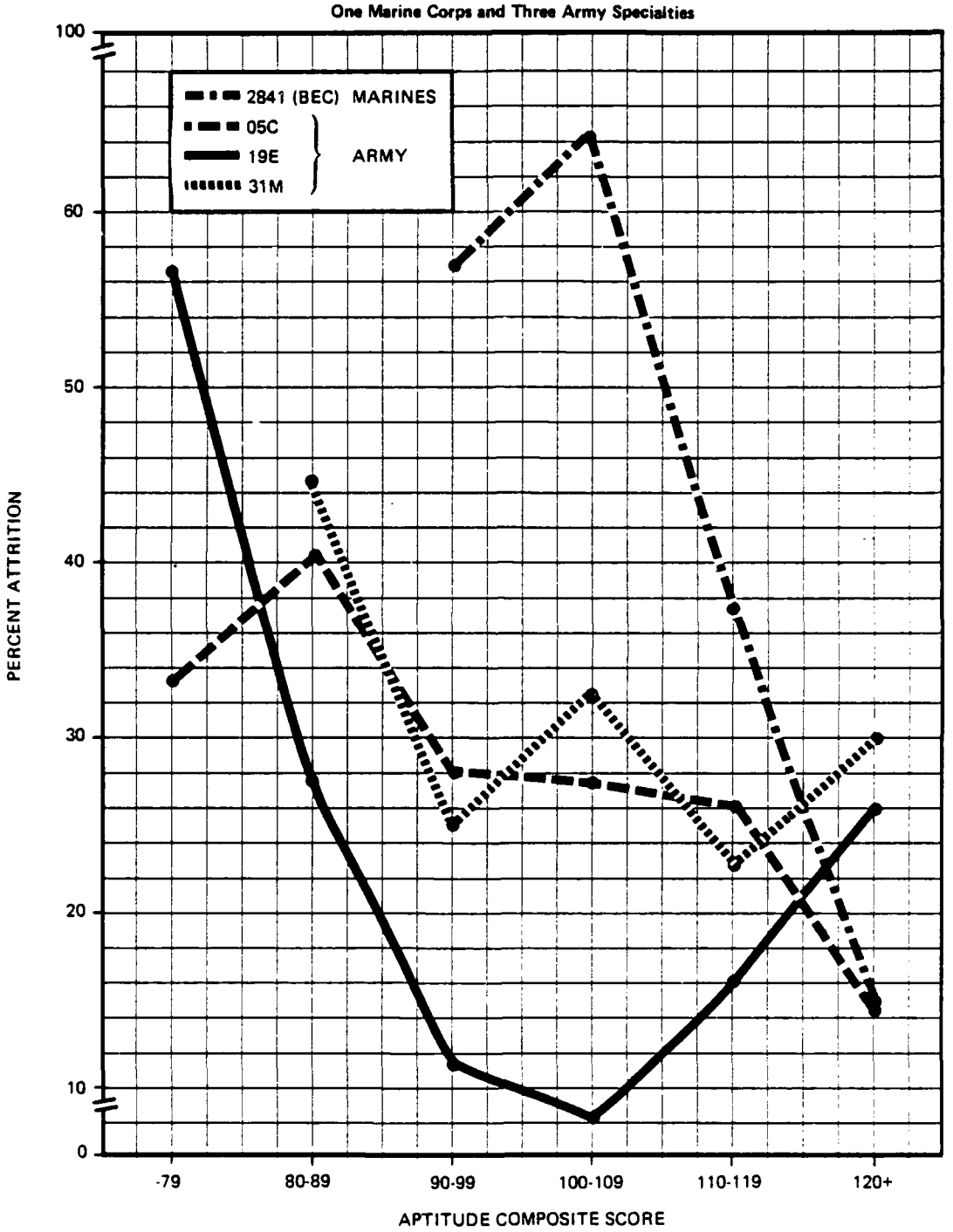


Figure 85
Time to Complete Training as a Function of AFQT Category
for One Marine Corps and Three Army Specialties

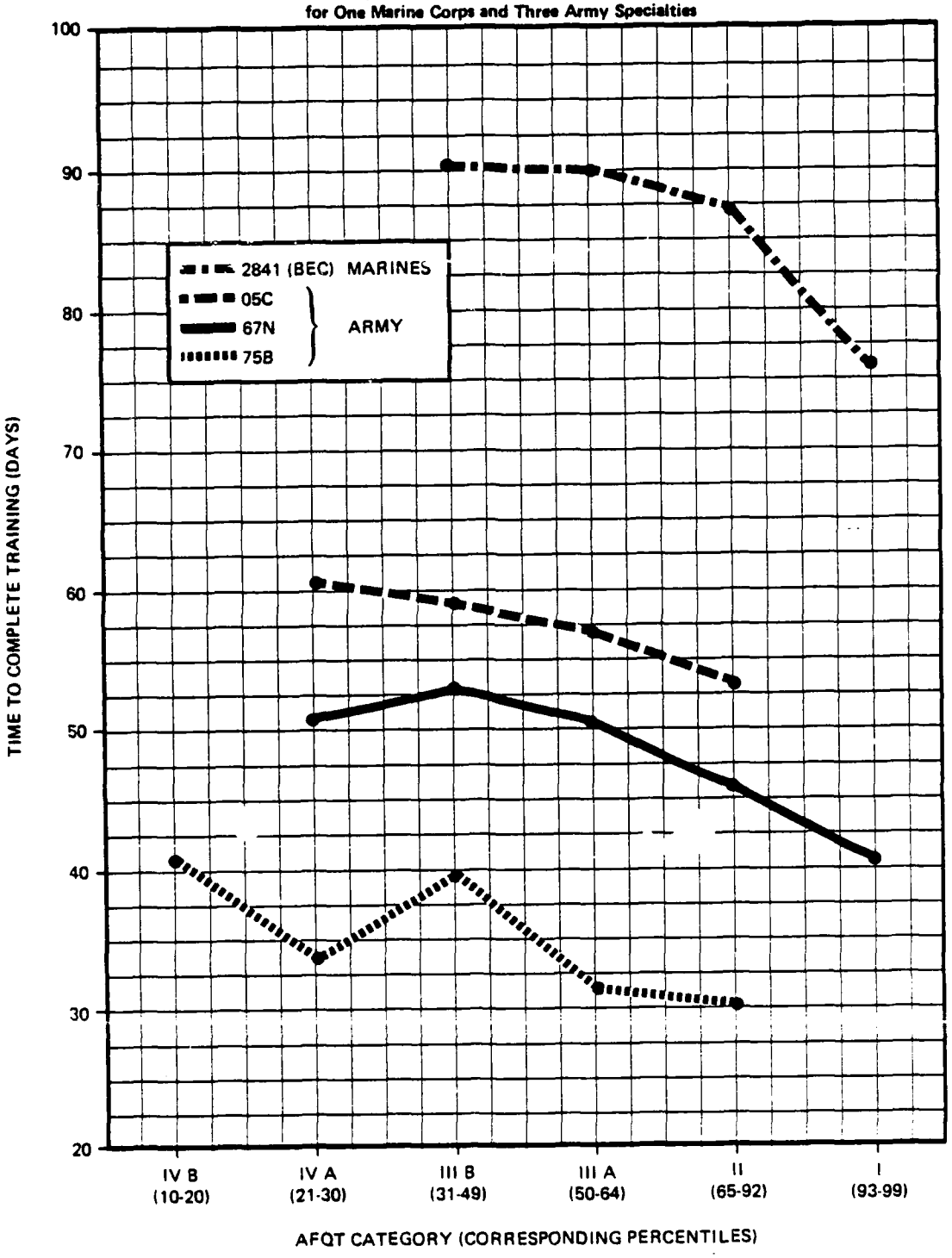


Figure 86
Time to Complete Training as a Function of Aptitude Composite Score for One Marine Corps and Three Army Specialties

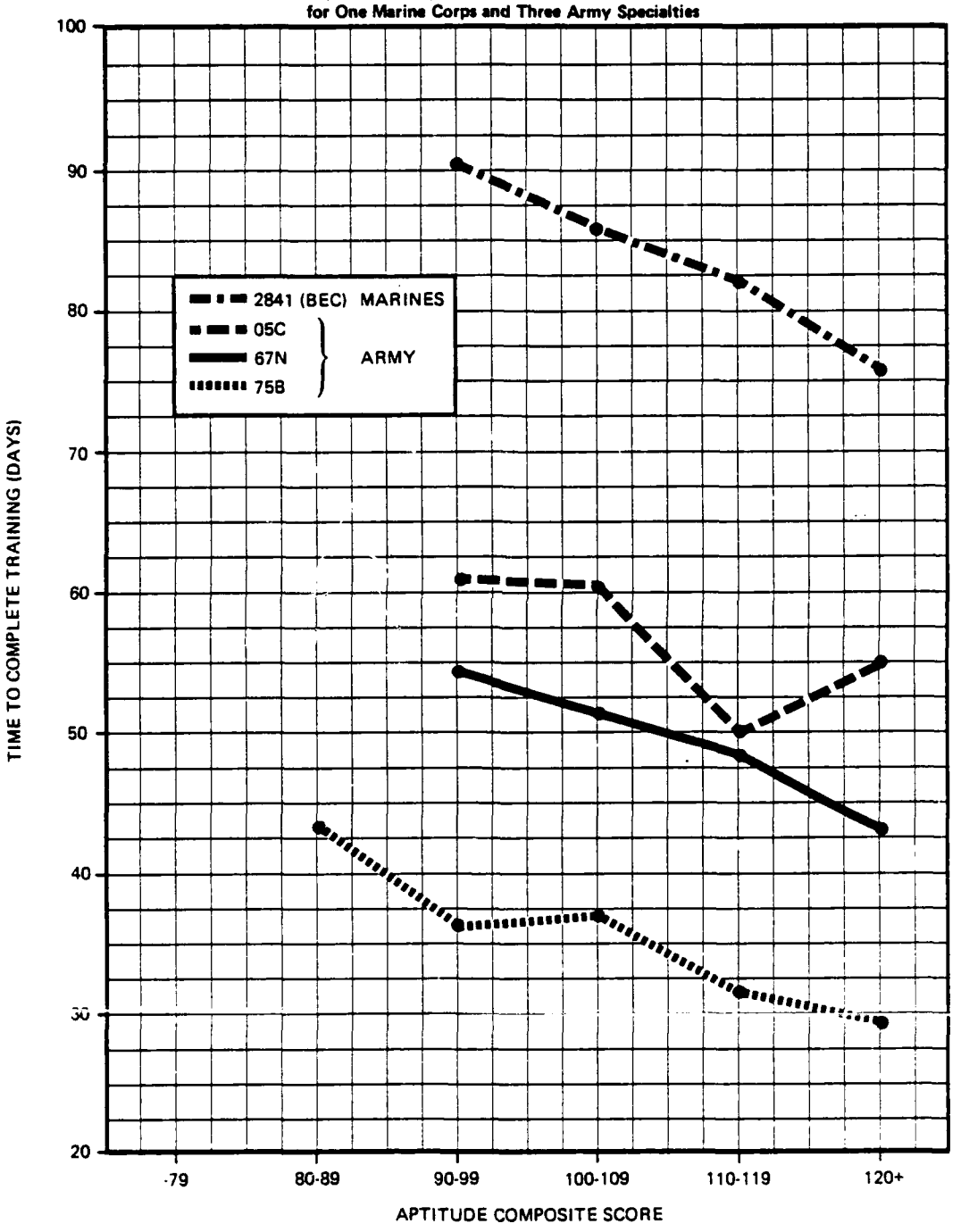


Figure 87
Mortar Qualification (MQ) Test Scores as a Function of AFQT Category and Education for MOS 11C (Indirect Fire Infantryman)

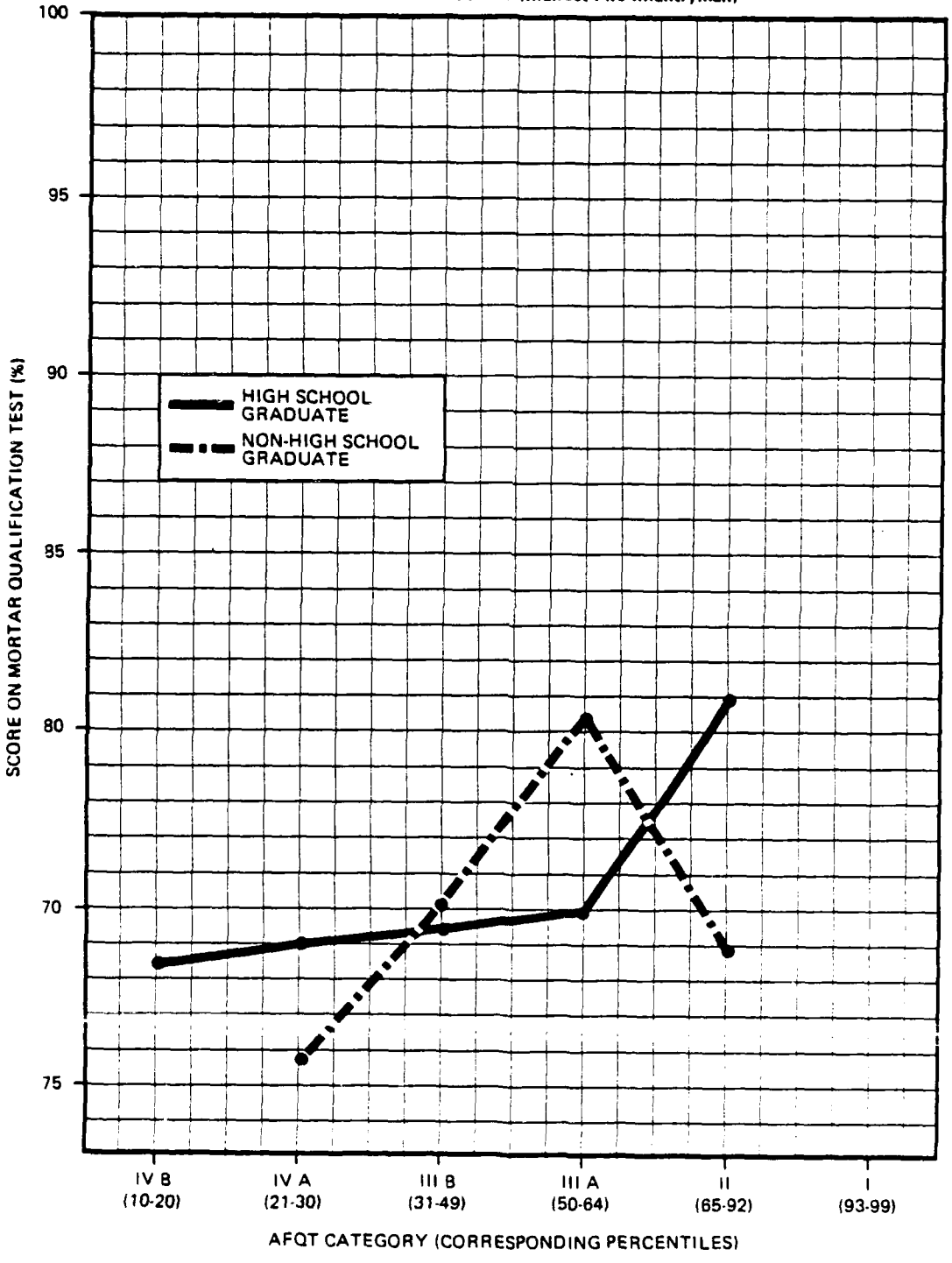


Figure 88
Mortar Qualification (MQ) Test Scores as a Function of Aptitude Composite Score and Education for MOS 11C (Indirect Fire Infantryman)

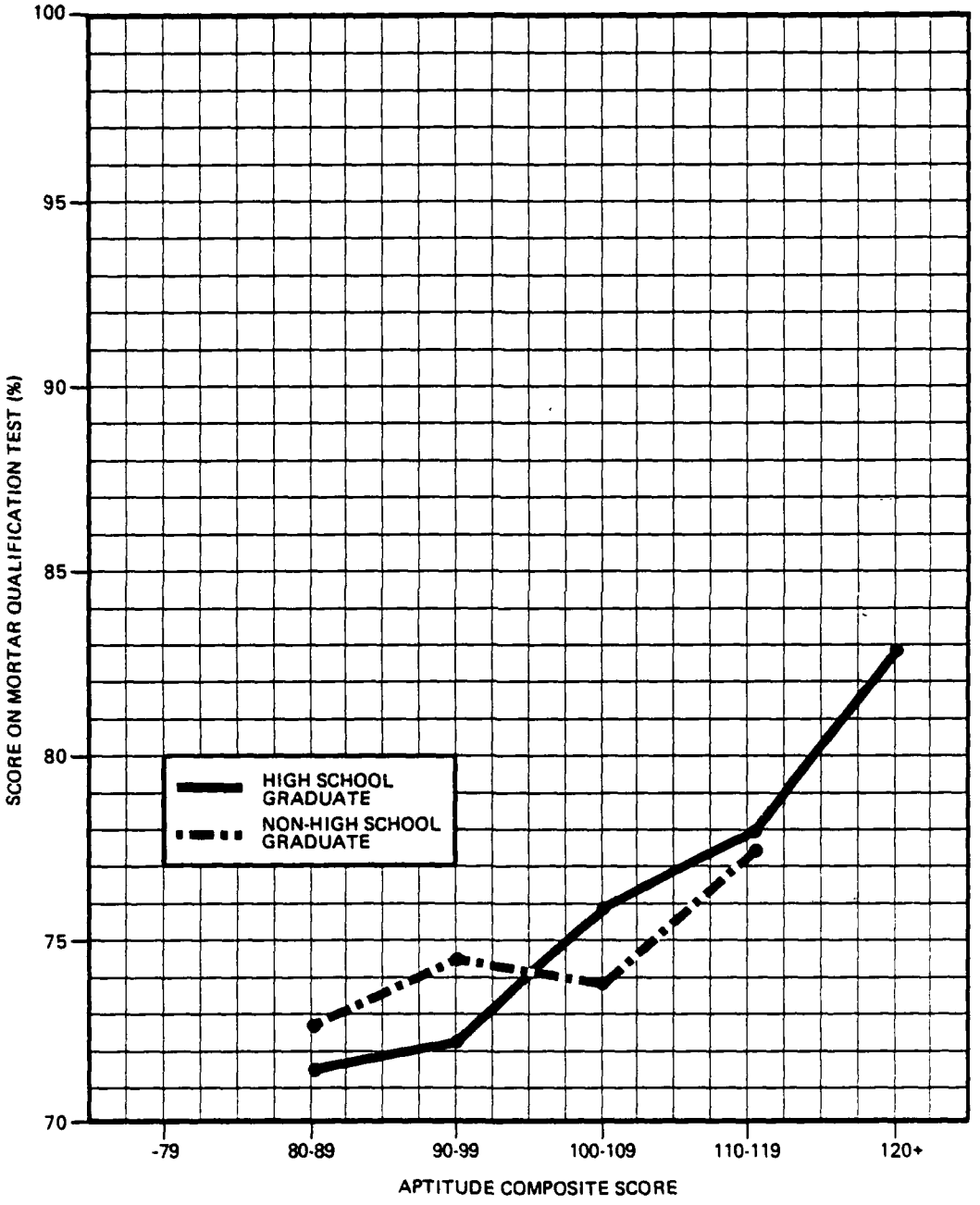


Figure 89
Peer Nomination as a Function of AFQT Category and Education
for Marine Corps Specialty 0311 (Infantryman)

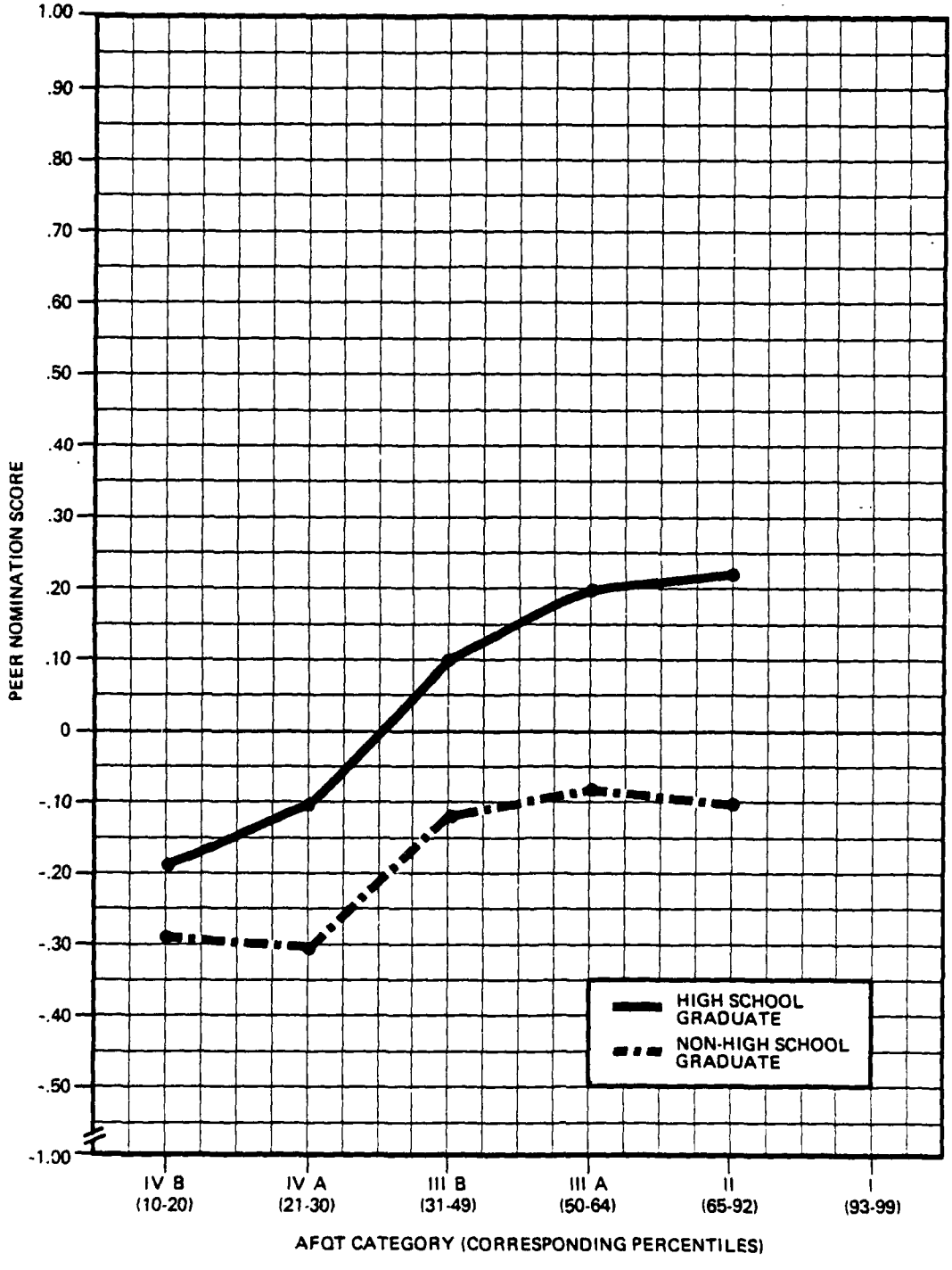


Figure 90
Peer Nomination as a Function of AFQT Category and Education
for Army MOS 11B (Infantryman)

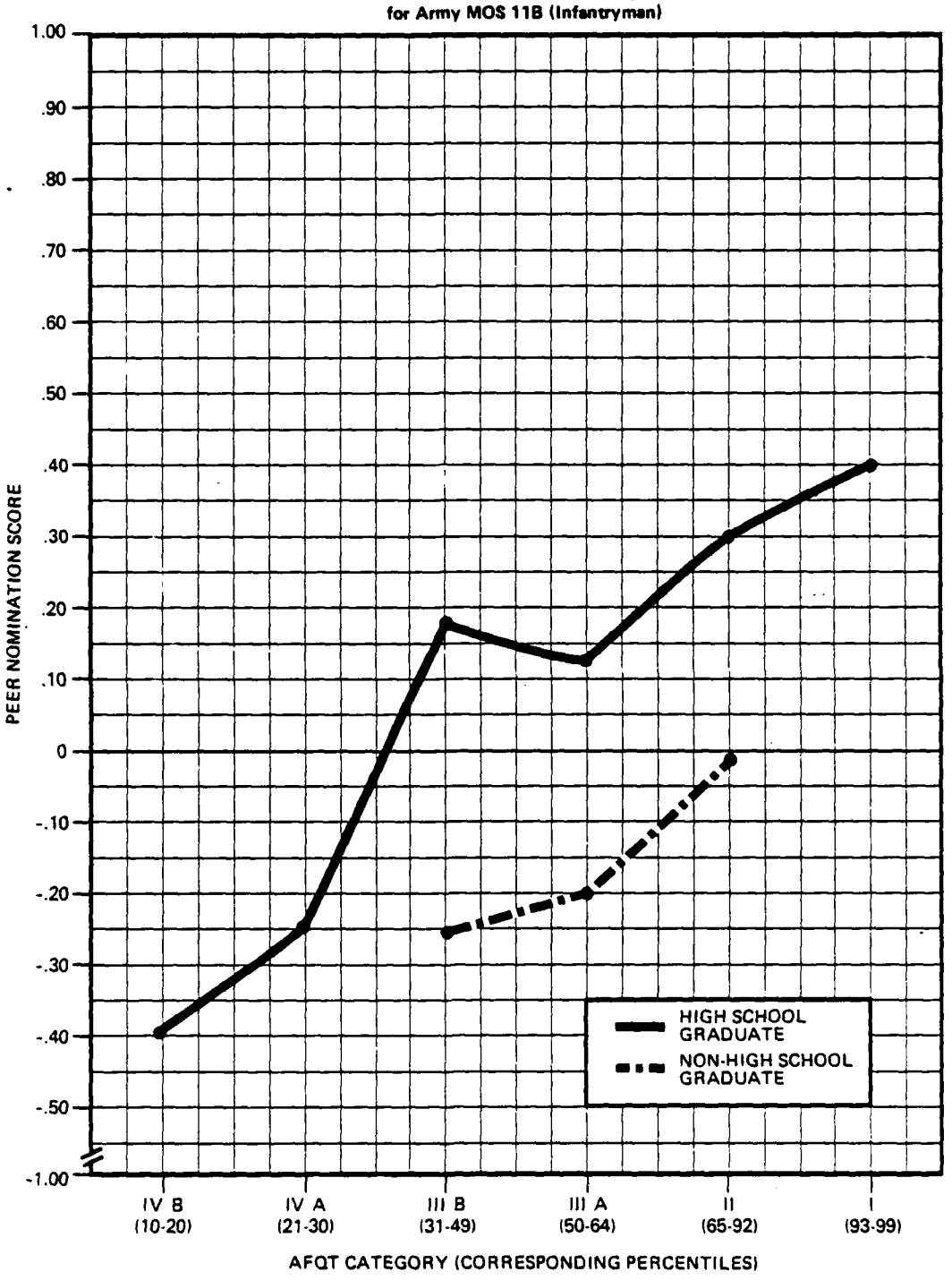


Figure 91
Peer Nomination as a Function of AFQT Category and Education for Army MOS 31M (Multichannel Communications Operator)

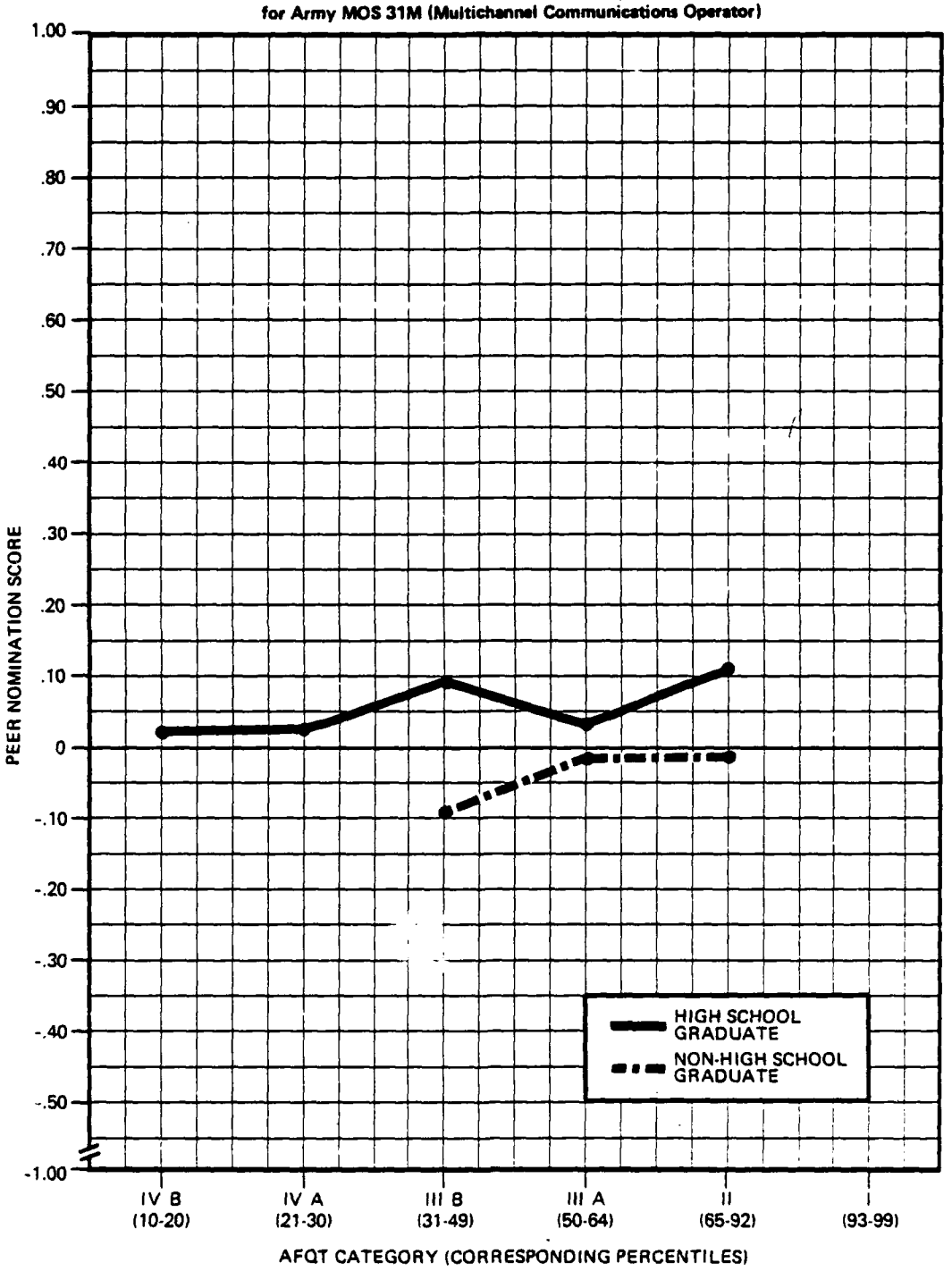


Figure 92
Peer Nomination as a Function of AFQT Category and Education
for Army MOS 73C (Finance Specialist)

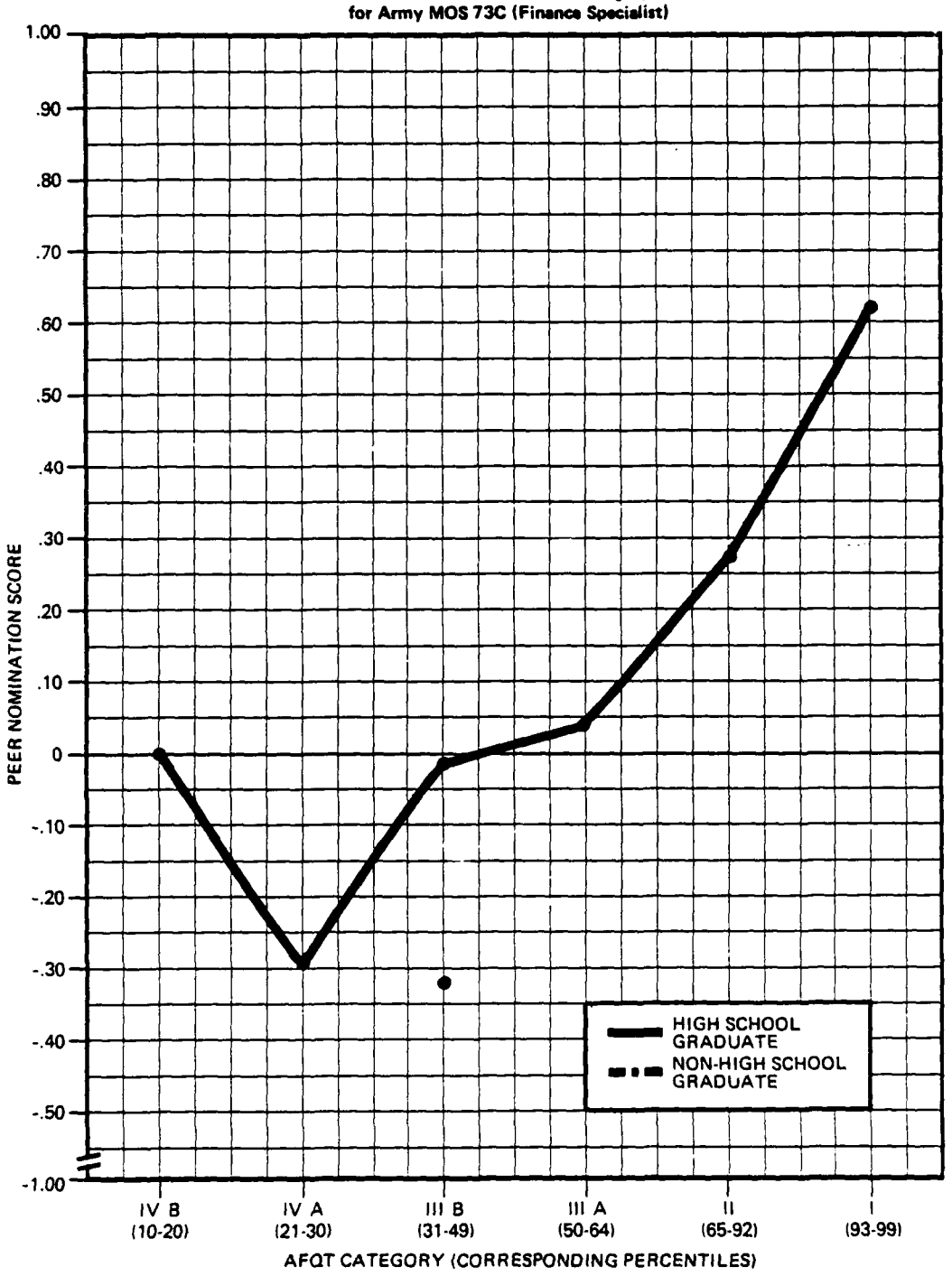


Figure 93
Peer Nomination as a Function of Aptitude Composite Score and Education for Marine Corpos Specialty 0311 (Infantryman)

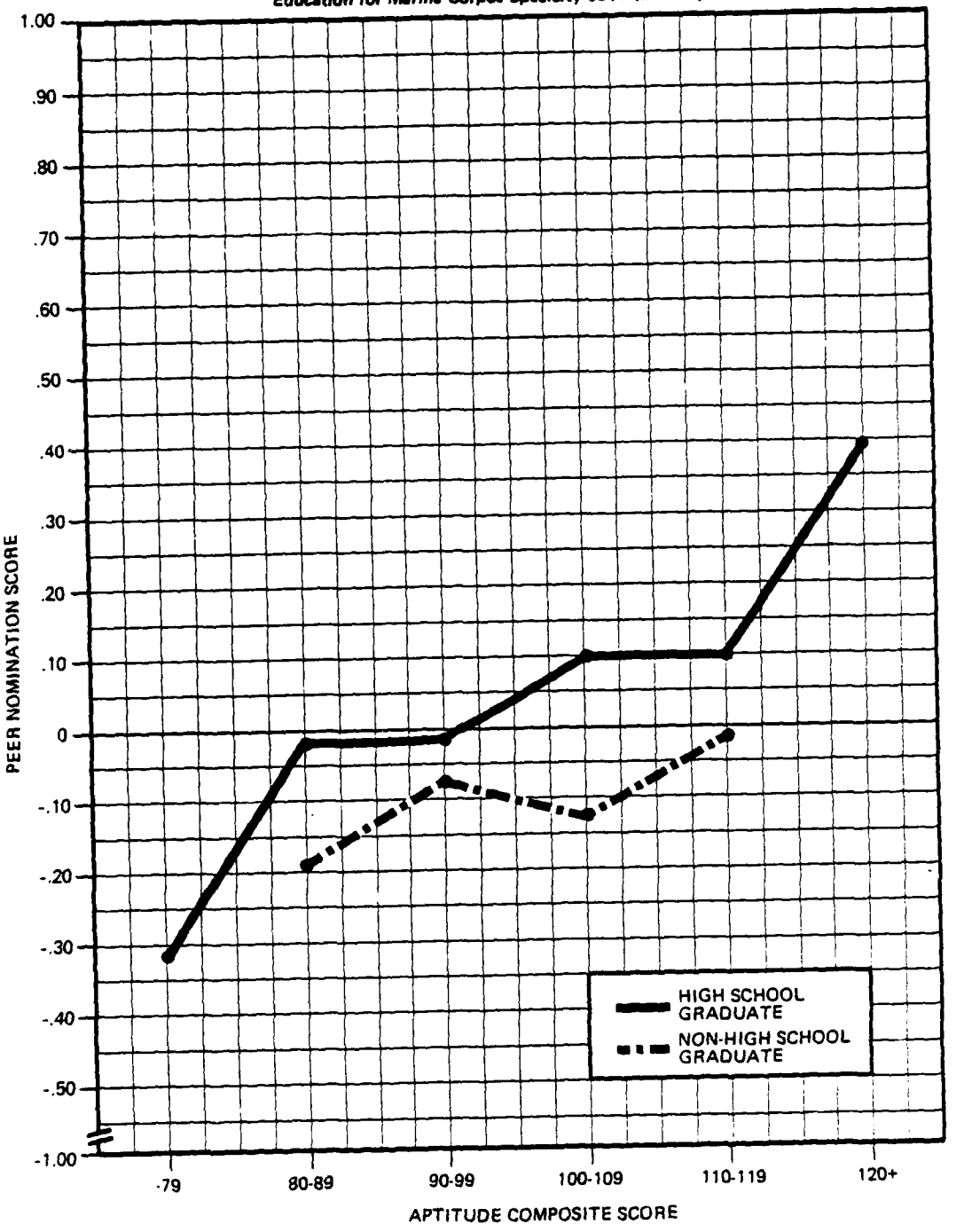


Figure 94
Peer Nomination as a Function of Aptitude Composite Score and Education for Army MOS 11B (Infantryman)

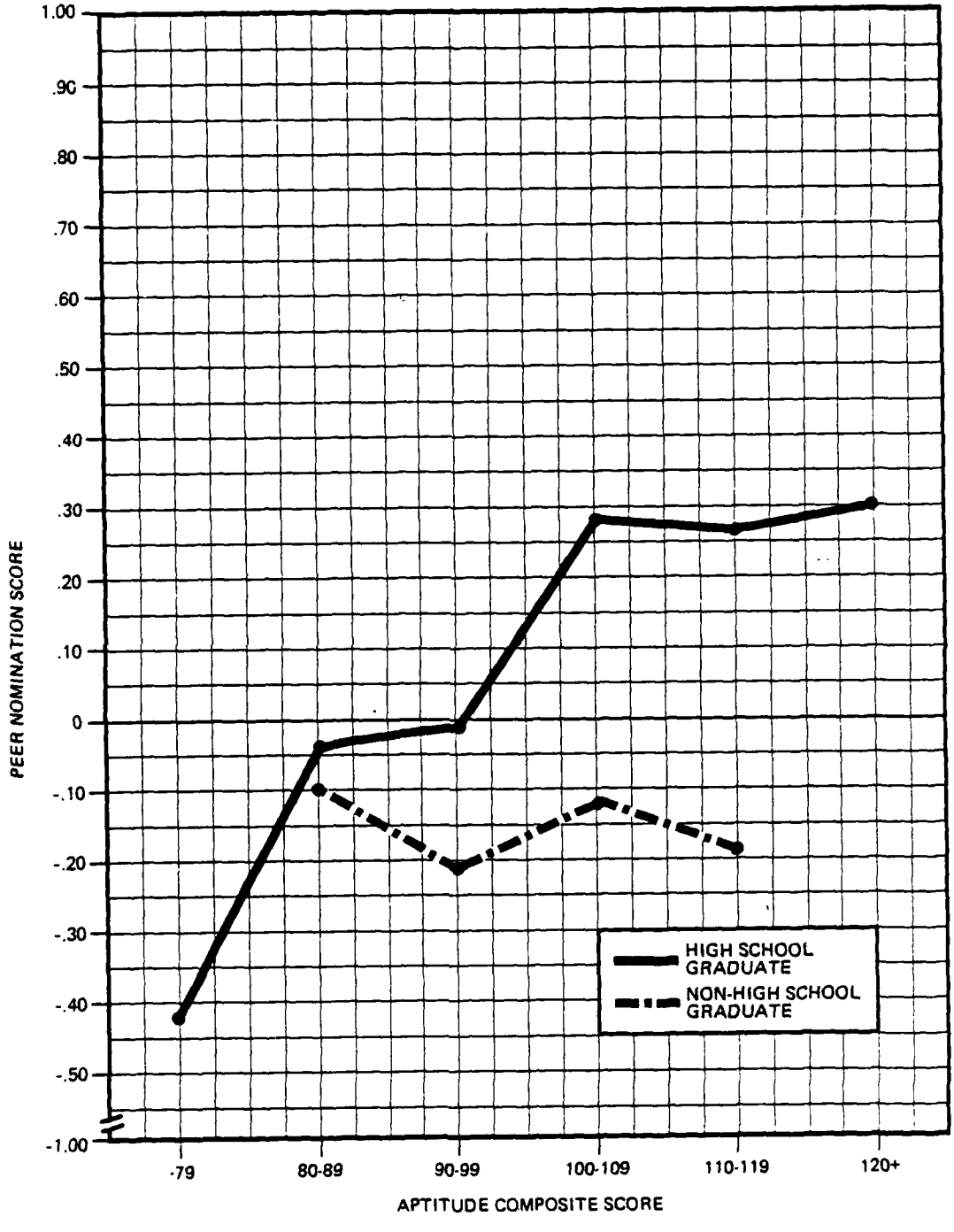


Figure 95
Peer Nomination as a Function of Aptitude Composite Score and Education for Army MOS 73C (Finance Specialist)

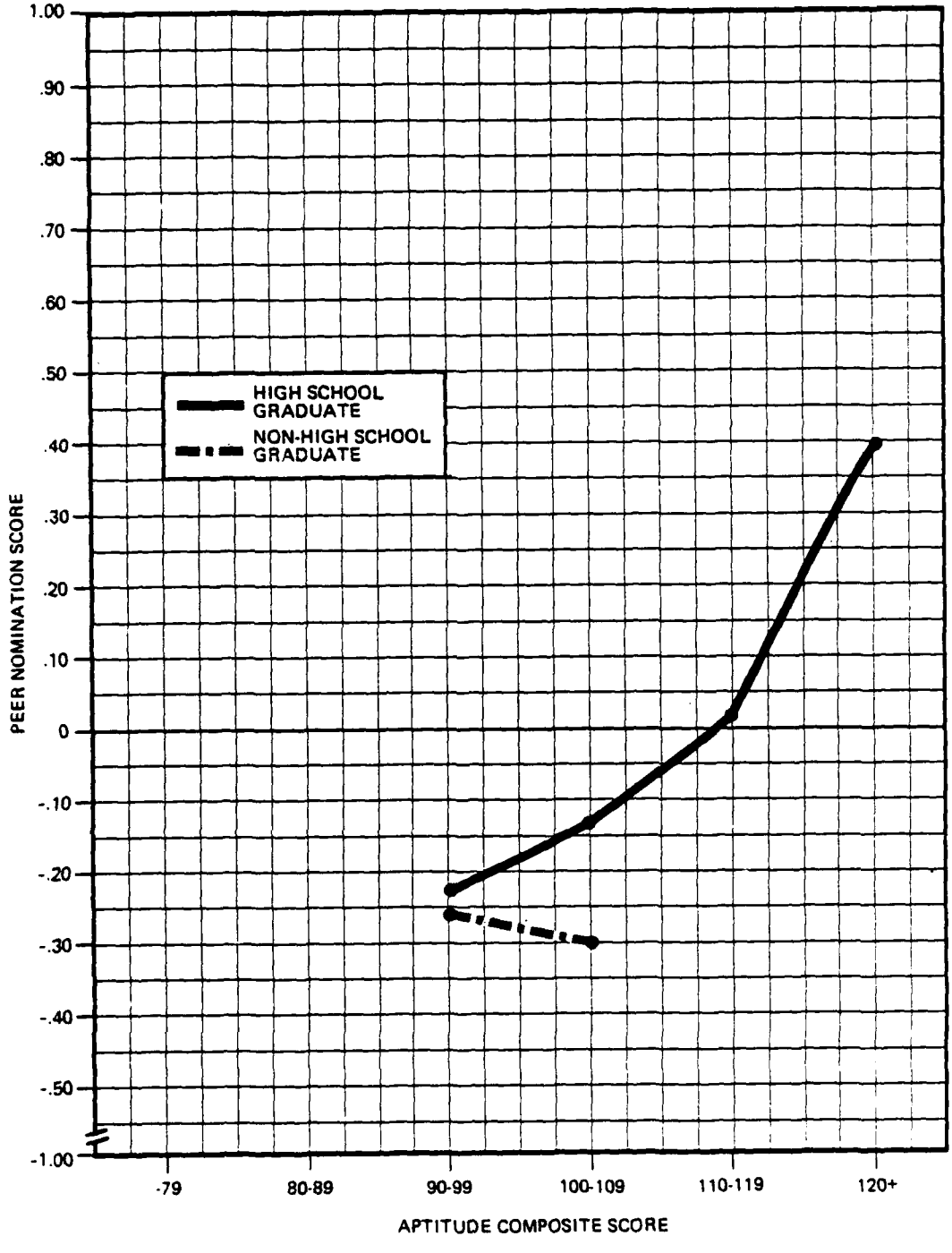


Figure 96
Peer Nomination as a Function of Aptitude Composite Score and Education for Army MOS 31M (Multichannel Communications Operator)

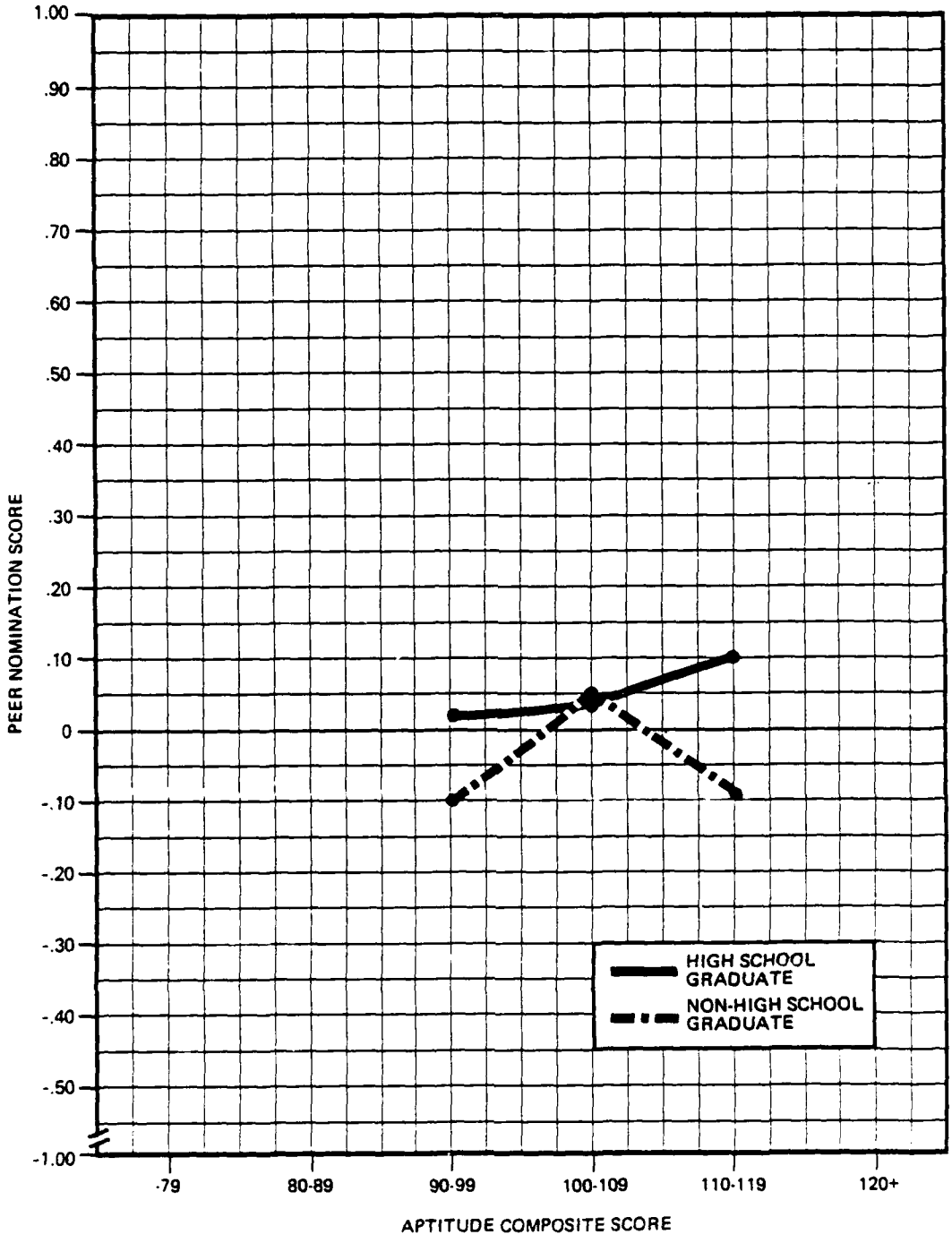


Figure 97
Instructor Rating Scores as a Function of AFQT Category and Education
for Army MOS 11B (Infantryman)

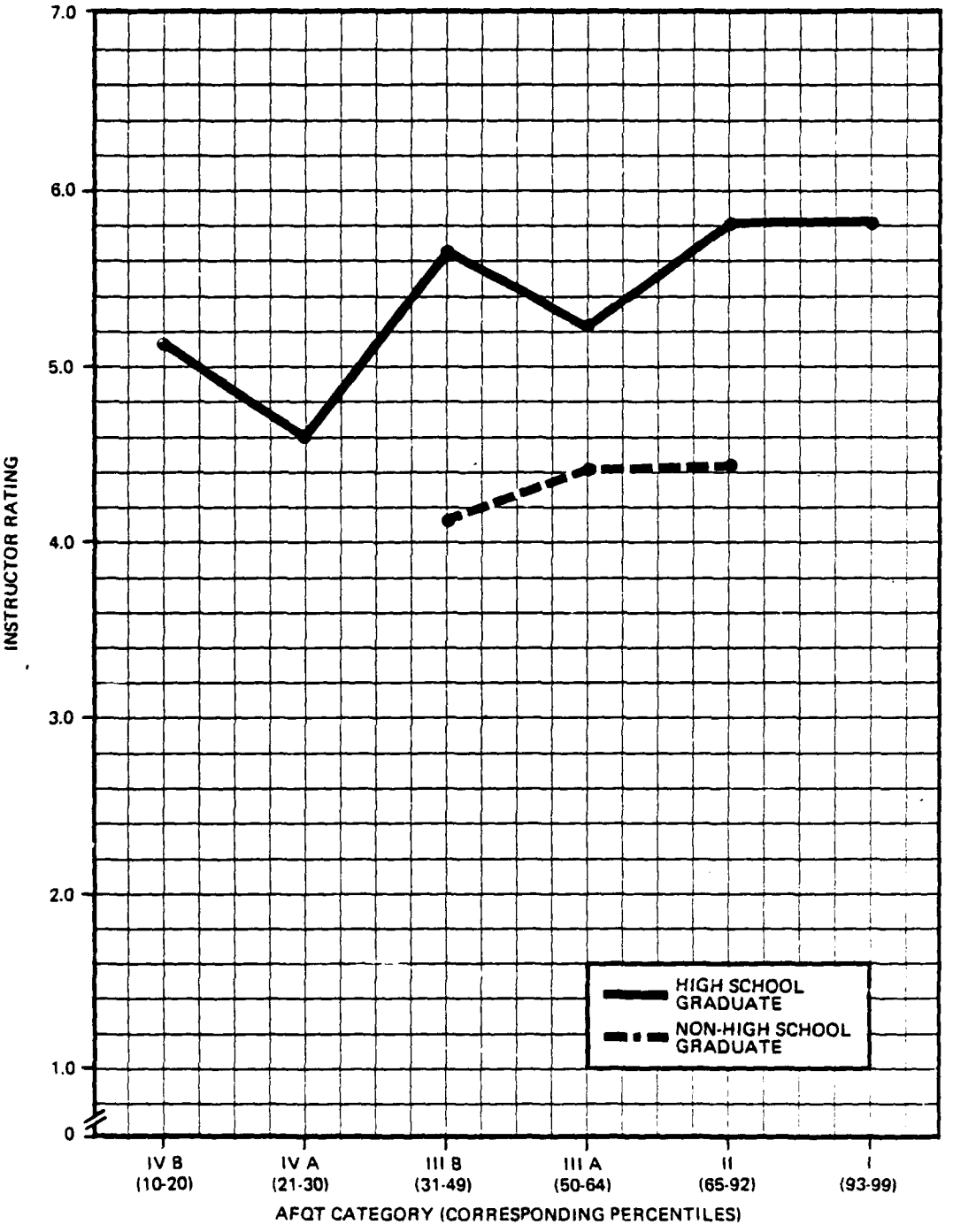
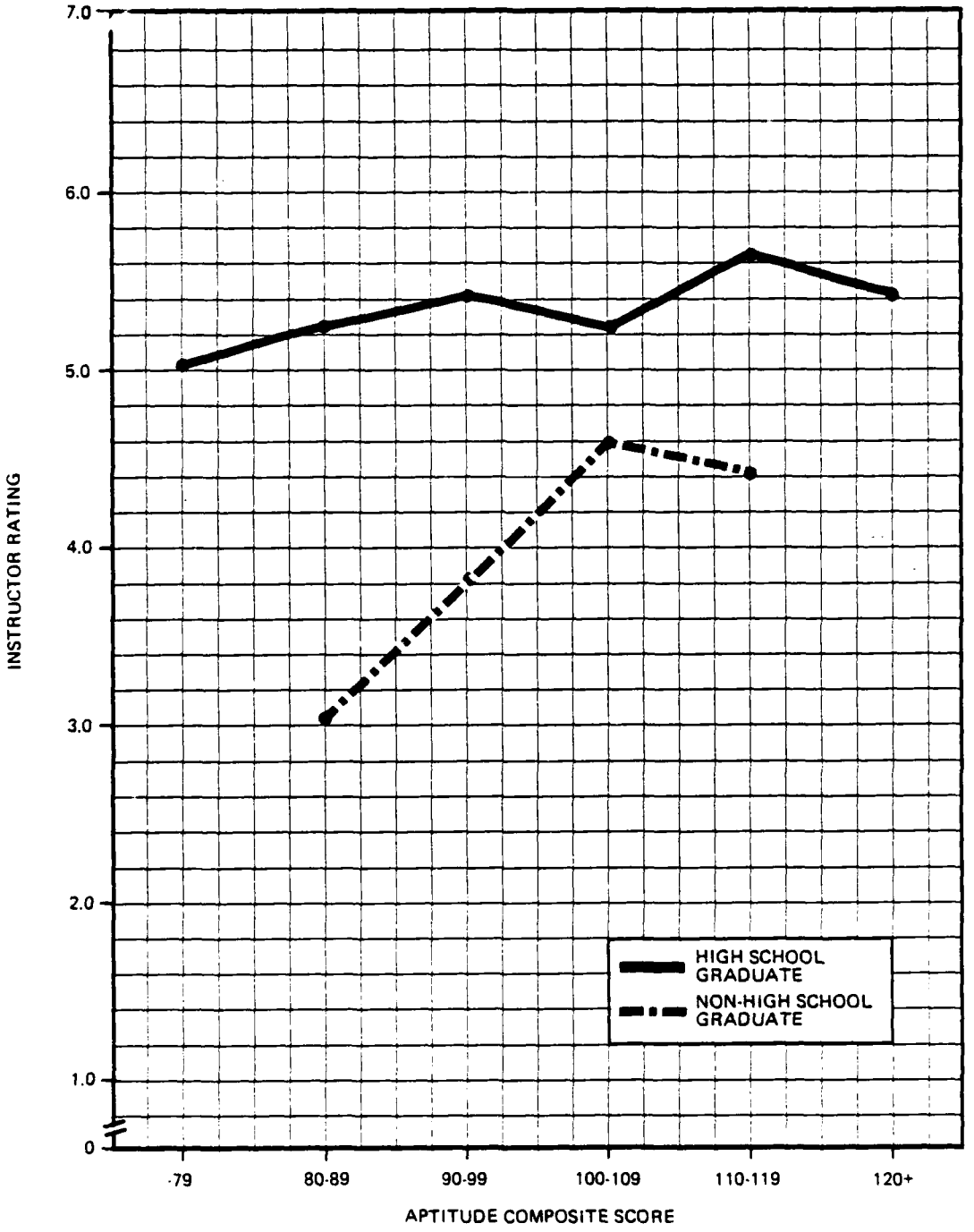


Figure 98
Instructor Rating Scores as a Function of Aptitude Composite Scores and Education for Army MOS 11B (Infantryman)



Appendix G

LITERATURE REVIEW

ISSUES RELATED TO SUBJECTIVE PERFORMANCE MEASURES DEVELOPED FOR THIS RESEARCH

Existing measures of enlistee training performance emphasize competence on specific tasks. Researchers agree, however, that successful performance in training and on the job is multidimensional, involving more than simply the ability to perform specified tasks (Uhlaner, Drucker & Camm, 1979; Fleishman, 1974; Stogdill, 1974; Helme, Willemin & Grafton, 1971). That is, soldier's attitudes, motivations and ability to communicate contribute to overall performance as well (Sellman & Silva, 1979; Massey, Mullins & Earles, 1978; Vineberg & Taylor, 1978). In fact, behavioral, attitudinal, and personality traits have been documented as valid and reliable predictors of successful job performance (Maier, 1973; Eastman & Leger, 1978; Downey & Duffy, 1978; Landy & Farr, 1980).

One of the goals of this research was to identify and develop alternative measures of performance in training. It was important that these measures be administratively feasible across jobs and across Services and have potential to contribute new information to performance measurement. Furthermore, these measures had to be developed within the time and resource constraints of the project. Consequently, rather than focusing on the development of either hands-on performance measures or performance-based tests, this research effort was directed towards more general aspects of performance.

Ratings

A variety of subjective methods may be employed to evaluate performance. The rating method is one of the most familiar and widely accepted subjective measures of individual effectiveness. Rating procedures consist of appraisals by raters of ratees on some set of attributes which can be expressed on some common quantitative scale. Ratings made by an individual or by a group may reflect important elements of personality that a grade or test score is likely to exclude (Uhlaner & Drucker, 1980). They are particularly useful in evaluating situations such as training where interaction with others is essential. Rating methods have widespread applicability as performance measures for military enlisted personnel.

In this project, two experimental rating measures of performance in training were developed. They were instructor ratings and peer nominations and are described in the following paragraphs.

Instructor Ratings

The instructor rating instrument required an instructor/supervisor in a training setting to rate each trainee in his class/squad on 10 dimensions according to a

seven-point scale. The criteria for administration included the requirement that an instructor be involved with a class on a continuous basis during training. However, at most schools instructors specialize and therefore do not see trainees in any one class long enough to become very familiar with them. This requirement, then, was met for only one MOS.

Peer Nominations

There are a variety of peer evaluation techniques, including ratings, rankings, high nominations, and full nominations. Nominations are obtained by asking each member of a group to select, for each of a series of attributions, a specified number of group members other than oneself, who match each attribution most closely. A full nomination technique, which requires choices on positive and negative attributions for each dimension or characteristic rated, is more readily accepted by raters than ratings or rankings, both of which require more difficult discriminations.

Peer nomination techniques, used in both military and industrial settings, have produced valid and reliable data. Reliabilities have typically been in the .70 to .90 range (Suci, Vallance & Glickman, 1954; Hollander, 1957; Fiske, 1960; Hammer, 1963; Flyer, 1964; Thomas, 1971; Shenk, Watson & Hazel, 1973; Downey, 1974; Mohr, 1975; Lewin & Zwany, 1976; Downey, Medland & Yates, 1976; Eastman & McMullen, 1976; Eastman & Leger, 1978; Downey & Duffy, 1978). Even the use of a paired-comparison peer evaluation technique does not significantly improve upon these reliabilities.

The procedure chosen for this study was the full nomination technique on six dimensions. The size of the peer groups ranged from 18 to 30. For each dimension, a rater chose six individuals as rating highest and six as lowest.

Considerations

Several considerations were taken into account in the selection and development of peer nominations and instructor ratings as experimental instruments for this project. While a number of rating procedures are capable of producing valid and reliable data, they are also susceptible, in varying degrees, to a number of types of measurement error and other factors which limit their utility as discriminative measures of job performance.

In developing rating procedures such as peer nominations and instructor ratings, it is important to address various types of rater errors such as halo, leniency, and central tendency, all of which threaten the discriminability of ratings.

- o A rater's judgment on one trait may influence his or her judgments on other traits, creating a halo effect. However, the rater may reduce this effect by rating all individuals on one behaviorally-defined trait before rating anyone on another, when he or she must rate more than one individual (Anastasi, 1979). Both the peer nomination and instructor rating methods employed in this study required raters to judge all individuals in their class/squad on one dimension before proceeding to the next dimension. This procedure was intended to reduce the halo effect.

- Organizational constraints on raters may cause inflation of ratings. This is called leniency error. Raters may be lenient in their evaluations when results are used for administrative purposes, or if they have to meet with ratees subsequently to discuss scores. It may be difficult for an instructor or supervisor who is held responsible for a trainee's level of proficiency to critically evaluate that individual's performance (Mullins & Ratcliff, 1979). Although in an operational setting, instructors might rate their trainees leniently, we attempted to control for leniency by telling supervisors that the principal purpose of these ratings was to validate AFQT/ASVAB rather than to serve as a criterion of performance.

- Raters may fail to use the entire rating scale, thereby committing errors of central tendency (Bergman & Siegel, 1972). Forced distribution and other order-of-merit procedures can reduce this error to some extent. The peer nomination technique, by requiring trainees to name the six highest and six lowest individuals in each dimension, reduced this error. However, an often cited weakness of the peer nomination technique is that it provides relatively less information about the middle of the distribution than about the extremes. Since each rater in this study chose six peers as possessing a positive attribution and six as possessing a negative attribution for each dimension rated, when the size of the peer group was 18, equal amounts of information were obtained regarding each third of the distribution. As the group size increases, however, proportionally less information is available about the middle of a distribution. This did not present a major problem in this study since group size for the peer nominations was limited to about 30.

When halo, leniency or central tendency errors are minimized, the effective size of the rating scale is increased, resulting in more discriminative ratings. However, a number of other factors may lead to unreliability of ratings.

- Raters who lack experience with subjective evaluation procedures tend to produce unreliable ratings. Research indicates, though, that rater accuracy can be significantly improved if raters can be trained in subjective measurement procedures (Bergman & Kujawski, 1969; Bergman & Siegel, 1972; Landy & Farr, 1980). However, in many cases, as in this study, such training is impractical. Our approach to this problem was to define the attributes or scale points in terms of behaviors; that is, to objectify the rated dimensions as much as possible.

- Raters do not always have the opportunity to view all job behavior relevant to a ratee's performance. It is important in a training setting that instructors have extensive contact with trainees over a period of several weeks to produce reliable ratings. Research indicates, however, that frequent peer association in a training situation for as short a period as eight weeks is sufficient for peers to make the required judgments accurately (Mohr, 1975), and where a peer group remains intact throughout training, reliable and valid peer evaluations can be obtained in as little as three to six weeks (Hollander, 1957). These guidelines were followed with the experimental measures introduced in this study by requiring that ratees were at least halfway through the training session

and that raters, whether peers or supervisors, had been with those they were rating throughout their training. In most cases, peers were in a position to observe a more typical sample of behavior than instructors or supervisors.

Characteristics of scaling techniques also contribute to measurement errors in the rating method. Therefore, when raters are making judgments on some qualitative scale, the scale must have certain properties to insure maximum accuracy.

- For example, the number of steps on the scale should be no more than raters can reliably differentiate but no fewer than required to make the necessary number of distinctions. The optimal number of steps have been variously estimated from five to nine (Maier, 1973; Mutell & Jacoby, 1972; Seashore, Indik, Georgopolos, 1962). The instructor ratings developed for this study required raters to differentiate between seven steps on the scale.

- For scales to be understood, they must have behavioral anchors or reference points relevant to persons the raters are evaluating. These anchors should be phrased in behavioral rather than in relative terms. Both of the experimental measures developed in this study used behavioral anchors. The verbal anchors for the instructor rating instrument were chosen so as to approximate equal intervals along the scale based on a summary of studies providing scale equivalents of verbal descriptors (Nystrom, 1976).

Rating format, administration, and scoring of rating evaluations are also important considerations.

- The ability of raters to be conscientious in their judgments varies inversely with the number of judgments they are required to make (Downey & Duffy, 1978). When the size of the rating group exceeds 20, the number of decisions gets excessive. Raters begin to suffer from fatigue and the reliability of their judgments declines. For this reason, both the instructor ratings and peer nominations were administered in groups with 30 or less individuals to be rated, with the average number about 24.

- The use of multiple evaluaters in peer nominations is likely to increase the validity of performance ratings (Karcher, Winer, Falk & Haggerty, 1952). There is greater agreement among multiple raters when they have had adequate opportunities to observe trainees on the job. Thus, because peers are generally able to perceive an individual in a wider range of situations than an instructor or supervisor, the resulting judgments tend to be more reliable.

Beyond the issues directly affecting rater accuracy, there are some issues relating to the properties of the numbers obtained in some rating procedures.

- The scaling properties of peer nomination data, while in fact ordinal, approximate interval data as the number in the evaluation group increases (Downey & Duffy, 1978).

- Job performance is multidimensional; thus, in order to develop the most useful criteria, it may be necessary to combine ratings on a number of dimensions to yield composite or profile scores (Blum & Naylor, 1968). The peer nomination and instructor rating data can be factor analyzed in order to identify these various performance dimensions.

- With instructor ratings, if raters use the whole scale and if the assumption can be made that different raters use the scale in the same way, then ratings of individuals can be compared within and across groups as well as to some standard of performance.

Summary

The innovative peer nomination and instructor rating measures developed for this research project may provide useful information when used in combination with existing devices to measure performance in training and on the job. That is, subjective measures may identify factors such as drive, persistence, and stability which may be related to how well a soldier uses the capabilities, aptitudes, or skills which are measured by existing tests (Uhlaner, 1970). In fact, measures such as peer nominations and instructor ratings, if they correlate with performance-based tests, may make an independent contribution to the prediction of potential success in training and on the job (Rundquist, Schneider & Frankfield, 1950; Kantor, Vitola, & Guinn, 1977). Research indicates that behavioral, attitudinal and personality traits can indeed be valid and reliable measures of successful job performance (Maier, 1973; Eastman & Lager, 1978; Downey & Duffy, 1978; Landy & Farr, 1980). Developmental, administrative, and scoring costs are significantly lower for subjective measures of job and training performance (such as instructor ratings and peer nominations) than for more objective performance measures. Therefore, if their validity and reliability can be documented, they deserve further consideration as potential cost-effective alternatives for measuring training performance.

REFERENCES

Anastasi, A. Fields of applied psychology (2nd ed). New York: McGraw-Hill, 1979.

Bergman, B. & Kujawski, C. How to rate your subordinates. Unpublished paper. The Atlantic-Richfield Company, 1969.

Bergman, B. A. & Siegel, A. J. Training evaluation and student achievement measurement: A review of the literature. (AFHRL-TR-72-3). Lowry AFB, Colorado: Air Force Human Resources Laboratory, January 1972.

Blum, M. & Naylor, J. Industrial psychology: Its Theoretical and social foundations. New York: Harper & Row, 1968.

Downey, R. G. Associate evaluations: Nominations vs. ratings. (ARI Technical Paper 253). Alexandria, Virginia: U. S. Army Research Institute for the Behavioral & Social Sciences, September 1974.

Downey, R. G., Medland, F. F., & Yates, L. G. Evaluation of a peer rating system for predicting subsequent promotion of senior military officers. Journal of Applied Psychology, 1976, 61, 206-209.

Downey, R. G. & Duffy, P. J. Review of peer evaluation research. (ARI Technical Paper 342). Alexandria, Virginia: Army Research Institute for the Behavioral and Social Sciences, October 1978.

Eastman, R. F. & McMullen, R. L. Reliability of associate ratings of performance potential by Army aviators. (ARI Research Memo-76-28). Alexandria, Virginia: U. S. Army Research Institute for the Behavioral and Social Sciences, November 1976.

Eastman, R. F. & Leger, M. Validity of associate ratings of performance potential by Army aviators. (ARI Research Memo-78-24). Alexandria, Virginia: U. S. Army Research Institute for the Behavioral and Social Sciences, October 1978.

Fiske, D. W. Variability among peer ratings in different situations. Educational and Psychological Measurement, 1960, 20, 283-290.

Fleishman, E. A. Twenty years of consideration and structure. In E.A. Fleishman, and J. G. Hunt (Eds.), Current Development in the Study of Leadership. Carbondale, Illinois: Southern Illinois University Press, 1974.

Flyer, E. Prediction by career field of first-term airmen performance from selection and basic training variables (PRL-TDR-64-5). Personnel Research Laboratory, March 1964.

Hammer, C. H. A simplified technique for evaluating basic trainees on leadership potential. (ARI Research Memo-63-10). Alexandria,

Virginia: U. S. Army Research Institute for the Behavioral and Social Sciences, 1963.

Helme, W. H., Willemin, L. P., & Grafton, F. C. Dimensions of leadership in a simulated combat situation. (ARI Technical Research Report 1172). Alexandria, Virginia: U.S. Army Research Institute for the Behavioral and Social Sciences, July 1971.

Hollander, E. P. The reliability of peer nominations under various conditions of administration. Journal of Applied Psychology, 1957, 41, 85-90.

Kantor, J. E., Vitola, B. M., & Guinn, N. Development and validation of the Air Force technical training student survey. (AFHRL-TR-77-27(I)). Brooks AFB, Texas: Air Force Human Resources Laboratory, June 1977.

Karcher, E. K., Jr., Winer, B. J., Falk, G. H., & Haggerty, H. R. A study of officer rating methodology: Validity and reliability of ratings by single raters and multiple raters. (Research Report 904). Alexandria, Virginia: U.S. Army Research Institute for the Behavioral and Social Sciences, April 1952.

Landy, F. J. & Farr, J. L. Performance rating. Psychological Bulletin, 1980, 87 (No. 1) 72-107.

Lewin, A. Y. & Zwany, A. Peer nominations: A model, literature critique and a paradigm for research. (Technical Report TR-1). Durham, North Carolina: Duke University Graduate School of Business Administration, February 1976.

Maier, N. R. F. Psychology in industrial organizations (4th ed.). Boston: Houghton Mifflin Company, 1973.

Massey, R. H., Mullins, C. J. & Earles, J. A. Performance appraisal ratings: The content issue. (AFHRL TR-78-69). Brooks AFB, Texas: Air Force Human Resources Laboratory, December 1978.

Mohr, E. S. Acceptability of associate ratings at branch basic schools. (ARI Technical Paper 268). Alexandria, Virginia: U. S. Army Research Institute for the Behavioral and Social Sciences, October 1975.

Mullins, C. J. & Ratliff, F. F. Criterion problems in criterion development for job performance evaluation: Proceedings from symposium 23 and 24 June 1977. (AFHRL TR-78-85). Brooks AFB, Texas: Air Force Human Resources Laboratory, February 1979.

Mutell, M. S. & Jacoby, J. Is there an optimal number of alternatives for Likert-scale items? Journal of Applied Psychology, 1972, 56, 506-509.

Nystrom, C. O. Questionnaire construction manual. Alexandria, Virginia: Fort Hood Field Unit. U.S. Army Research Institute for the Behavioral and Social Sciences, September 1976.

Rundquist, E. A., Schneider, D. E. & Frankfeld, E. Development of an enlisted efficiency report. (Research Note 51-5). U.S. Army Adjutant General's Office, May 1950.

Seashore, S.E., Indik, B. P. & Georgopoulos, B.S. Relationships among criteria of job performance. Journal of Applied Psychology, 1962, 44, 195-202.

Sellman, W. S. & Silva, W. T. The criterion problem: A personnel management perspective. In Mullins, C. J. & Ratliff, F. F. Criterion problems in criterion development for job performance evaluation: Proceedings from symposium 23 and 24 June 1977. (AFHRL TR-78-85). Brooks AFB, Texas: Air Force Human Resources Laboratory, February 1979.

Shenk, F., Watson, T. W., Hazel, J. T. Relationship between personality traits and officer performance and retention criteria. (AFHRL-TR-73-4). Brooks AFB, Texas: Air Force Human Resources Laboratory, May 1973.

Stogdill, R. M. Handbook of leadership: A survey of theory and research. New York: The Free Press, 1974.

Suci, G. J., Vallance, T. R. & Glickman, A. S. An analysis of peer ratings. (Technical Bulletin No. 54-9). Newport, Rhode Island: Bureau of Naval Personnel, September 1954.

Thomas, P. J. An evaluation of methods for predicting job performance of personnel. (Technical Bulletin STB-72-4). San Diego California: Navy Personnel and Training Research Laboratory, September 1971.

Uhlaner, J. E. Human performance jobs and systems psychology — the system measurement bed. (Technical Report S-2). Alexandria, Virginia: U.S. Army Research Institute for the Behavioral and Social Sciences, October 1970.

Uhlaner, J. E. & Drucker, A. J. Military research on performance criteria: A change of emphasis. Human Factors, 1980, 22, 131-139.

Uhlaner, J. E., Drucker, A. J. & Camm, W. B. Army research and the criteria problem: A change of emphasis. In Mullins, C. J. & Ratliff, F. F. Criterion problems in criterion development for job performance evaluation: Proceedings from symposium 23 and 24 June 1977. (AFHRL TR-78-85). Brooks AFB, Texas: Air Force Human Resources Laboratory, February 1979.

Vineberg, R. & Taylor, E. N. Alternatives to performance testing: Tests of task knowledge and ratings. (HumRRO Professional Paper 6-78). Alexandria, Virginia: Human Resources Research Organization, March 1978.

Appendix H

EXPERIMENTAL TRAINING PERFORMANCE MEASURES

Part I: Specific Criteria and Procedures for Administering the Peer Nomination and Instructor Ratings

Part II: Experimental Training Performance Measurement Instruments

- o Peer Nomination Form
- o Instructor Rating Form

SPECIFIC CRITERIA AND PROCEDURES FOR ADMINISTERING THE PEER NOMINATIONS AND INSTRUCTOR RATINGS

Requirements

One requirement for administration of the peer nominations was that enlistees had to have been at least half of the way through their training courses. Additionally, peer nominations were only administered in lockstep courses or in courses in which trainees had ample opportunity to observe each other's work. Group size was important as well. Peer nominations were administered in MOS training groups with between 18 and 30 members. Four MOS training courses which met these requirements were Marine Corps 0311 (Infantryman), Army 11B (Infantryman), Army 31M (Multichannel Communications Operator), and Army 73C (Finance Specialist).

In addition to the two criteria for administering peer nominations, two other conditions were required for administering the instructor ratings. The instructor ratings were only administered to training groups with 25 or fewer trainees. It was also required that an instructor teach trainees throughout training. At many schools instructors teach only a limited portion of a course and therefore do not observe trainees in any one class long enough to become very familiar with them. This requirement was met by only one MOS, 11B (Infantryman).

Procedures

The administration of peer nominations required trainees to assemble in a group, where they were provided with numbered rosters. Each trainee was asked to draw a line through their own name and roster number. Trainees were then asked to write down in the spaces provided on the peer rating form the roster number of six unit peers who best fit each of 12 attributional statements (along six dimensions) pertaining to performance in training. Each of six dimensions (motivation, ability to communicate, leadership ability, proficiency with equipment, cooperativeness, and overall soldiering) was stated in both a positive and negative manner (e.g., "Works well with most equipment" and "Has difficulty working with equipment") so that trainees identified the six "best" and "worst" peers on each of the six dimensions.

The procedure for administering the instructor ratings was as follows. Supervisors/instructors were asked to assign each traineee in their group a rating on a scale from 1 to 7 on each of 10 performance-related dimensions (i.e., dependability, written and oral communication, motivation, leadership, attitude toward others, confidence, attitude towards supervision, ability to work with equipment, organizational ability and predicted job performance). Supervisors were told to make sure that each soldier had been given a rating on a dimension before moving on to the next dimension. To assist the instructor in making accurate ratings, behavioral anchors (or descriptions of different levels) of the dimension (e.g., lacks motivation, seldom tries to succeed in training vs. highly motivated, tries hard to succeed in training) were provided at key points along each scale. Supervisors took about one hour to complete their ratings.

PEER NOMINATION FORM

NAME_____ DATE_____

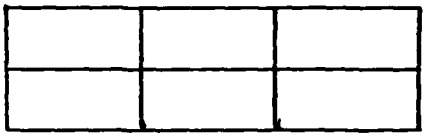
SOCIAL SECURITY NUMBER_____

This is a questionnaire which you will use to rate the trainees in your unit. All ratings will be kept private. Since we are just trying out this type of rating procedure, these ratings will not influence your grades. However, it is important that you fill this form out carefully.

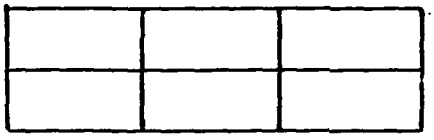
A copy of the roster for your unit is attached. Look at it. You will use it to make your ratings. Notice that a number is printed next to each name. Since you will not be rating yourself, draw a line through your name and roster number. Do it now.

Now look at the first statement below: "Highly motivated, tries hard to succeed in training." Underneath this statement there are six boxes. Your task is to look at the roster and choose the six trainees who are most like this statement, and write their roster numbers in the boxes. Continue reading the statements to find the six trainees who are most like each statement. When you are finished, check to make sure you have filled in all of the boxes.

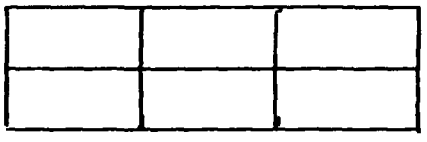
1. Highly motivated, tries hard to succeed in training.



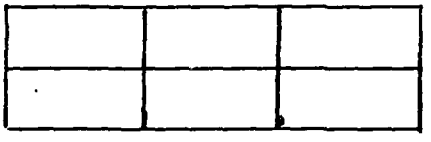
2. Lacks motivation, seldom tries hard to succeed in training.



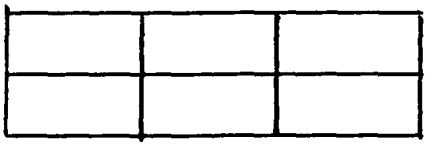
3. Communicates well, explanations are understandable and well organized.



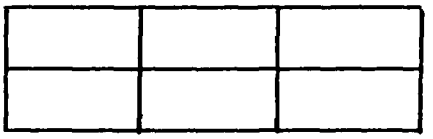
4. Communicates poorly, explanations are difficult to understand and disorganized.



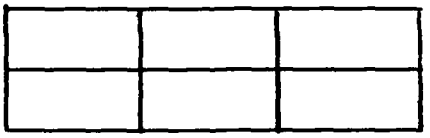
5. Eager to take charge, knows what needs to be done.



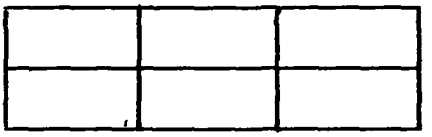
6. Cannot be counted on to take charge, seldom knows what needs to be done.



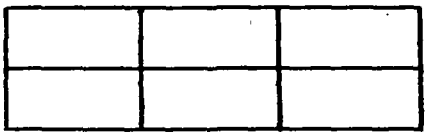
7. Works well with most equipment.



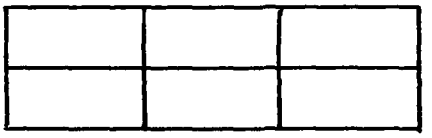
8. Has difficulty working with equipment.



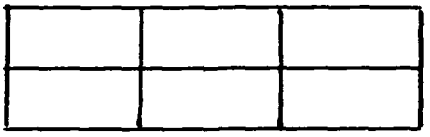
9. Very cooperative, works well with others.



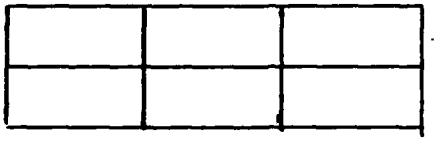
10. Seldom cooperative, unable to work with others.



11. Most likely to make a good infantryman.



12. Least likely to make a good infantryman.



INSTRUCTOR RATING FORM

DATE_____

INSTRUCTOR NAME_____ COURSE (MOS)_____

HOW MANY WEEKS HAVE YOU BEEN IN CONTACT WITH TRAINEES?_____

INSTRUCTOR/FACILITATOR RATING OF TRAINEE PERFORMANCE

This is an evaluation form which you will use to rate trainees you have dealt with. These ratings will be used for research purposes only and will be kept confidential. The purpose of this research is to try to improve the quality of recruits entering the Army, and your sincere evaluations are essential. On the next page you will be asked to rate a group of trainees in terms of their "dependability." You will rate each trainee on a scale from 1, the least dependable, to 7, the most dependable. To assist you in making accurate ratings, descriptions of different levels of dependability are provided at key points along the scale. For each trainee circle the most appropriate number from 1 through 7. When you have completed each page, make sure you have assigned a rating to each trainee.

DEPENDABILITY

	Seldom dependable, requires constant supervision		Sometimes dependable, requires considerable supervision		Fairly dependable, requires some supervision		Dependable most of the time, requires little supervision
(Soldiers' Names)	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

COMMUNICATION: WRITTEN AND ORAL

	Explanations are difficult to understand and dis-organized		Explanations are somewhat understandable but poorly organized		Explanations are fairly understandable and adequately organized		Explanations are very understandable and well organized
(Soldiers' Names)	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

MOTIVATION

(Soldiers' Names)	Lacks motivation, seldom tries to succeed in training		Occasionally motivated, sometimes tries to succeed in training		Usually motivated, generally tries to succeed in training		Highly motivated, tries hard to succeed in training
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

LEADERSHIP

	Cannot be counted on to take charge, seldom knows what needs to be done		Must be asked to take charge, sometimes knows what needs to be done		Willing to take charge, usually knows what needs to be done		Eager to take charge, knows what needs to be done
(Soldiers' Names)	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

ATTITUDE TOWARDS OTHERS

	Seldom cooperative, unable to work with others		Sometimes cooperative, occasionally able to work with others		Fairly cooperative, is able to work with others		Very cooperative, works well with others
(Soldiers' Names)	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

CONFIDENCE

	Seldom appears confident		Occasionally appears confident		Usually appears confident		Almost always appears confident
(Soldiers' Names)	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

ATTITUDE TOWARDS SUPERVISION

	Resents supervision, seldom willing to listen to instructions		Sometimes accepts supervision, sometimes willing to listen to instructions		Generally accepts supervision, usually willing to listen to instructions		Appreciates supervision, eager to listen to instructions
(Soldiers' Names)	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

ABILITY TO WORK WITH EQUIPMENT

	Has difficulty working with equipment		Works adequately with some equipment		Works adequately with most equipment		Works well with most equipment
(Soldiers' Names)	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

ORGANIZATIONAL ABILITY

	Generally disorganized		Occasionally appears organized		Usually appears organized		Appears well organized most of the time
(Soldiers' Names)	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

PREDICTED JOB PERFORMANCE

(Soldiers' Names)	Most likely make an unsatisfactory radio teletype operator		Most likely make a barely adequate radio teletype operator		Most likely make a fairly good radio teletype operator		Most likely make a very good radio teletype operator
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

Appendix I

UNCORRECTED CORRELATIONS BETWEEN AFQT/APTITUDE COMPOSITE SCORES AND TRAINING PERFORMANCE MEASURES

This Appendix contains Tables 28 through 31. These tables display the uncorrected correlations between AFQT and aptitude composite scores and measures of training performance. They correspond to Tables 14 through 17 in Section III, in which coefficients are corrected for restriction of range.

Table 28

Uncorrected Correlations Between AFQT Scores and Measures of Training Performance

MOS	Final Course Grade	Attrition*	Time-to-Complete	Peer Nom.	Instructor Rating	Alternate Performance Measures
0311 Infantryman	.20			.35		
11B Infantryman	.23			.44	.22	
11C Indirect Fire Infantryman	.10					.18
19E - Armor Crewman	.35	.09				
05C - Radio Teletypewriter Operator	.27	.16	-.22			
31M - Multichannel Communications Operator	.31	.05		.11		
2841 - Basic Electronics	.34	.42	-.15			
2841 - Radio Fundamentals	.29					
2841 - Ground Radio Repair	.25					
67N - Utility Helicopter Repairer	.64		-.44			
73C - Finance Specialist	.38			.57		
75B - Personnel Administration Specialist	.32		-.29			

* Because attrition is a dichotomous variable, the resulting biserial correlation coefficients cannot be translated to Pearson product-moment coefficients. Thus, correlations shown underestimate the relationship between AFQT scores and attrition.

Table 29

Uncorrected Correlations Between Aptitude Composite Scores and Measures of Training Performance

MOS	Final Course Grade	Attrition*	Time-to-Complete	Peer Nom.	Instructor Rating	Alternate Performance Measures
0311 Infantryman	.28			.36		
11B Infantryman	.22			.33	.21	
11C Indirect Fire Infantryman	.16					.24
19E - Armor Crewman	.39	.14				
05C - Radio Teletypewriter Operator	.27	.10	-.24			
31M Multichannel Communications Operator	.44	.18		.19		
2841 - Basic Electronics	.51	.39	-.34			
2841 - Radio Fundamentals	.43					
2841 - Ground Radio Repair	.22					
67N - Utility Helicopter Repairer	.60		-.42			
73C - Finance Specialist	.43			.63		
75B - Personnel Administration Specialist	.28		-.31			

*Because attrition is a dichotomous variable, the resulting biserial correlation coefficients cannot be translated to Pearson product-moment coefficients. Thus, correlations shown underestimate the relationship between selector composite and attrition scores.

Table 30

Uncorrected Correlations Between AFQT Scores and Measures of Training Performance Across Racial/Ethnic Groups for Four MOSs

		Final Course Grade			Attrition*		
	MOS	White	Black	Hispanic	White	Black	Hispanic
05C	Radio Teletypewriter Operator	.19	.22	**	.29	.03	**
31M	Multichannel Communications Operator	.45	.03	**	.16	-.02	**
73C	Finance Specialist	.35	.33	**			
75B	Personnel Administration Specialist	.34	.25	**			

		Time-to-Complete			Peer Nomination		
		White	Black	Hispanic	White	Black	Hispanic
05C	Radio Teletypewriter Operator	.00	-.31	**			
31M	Multichannel Communications Operator				.20	-.16	**
73C	Finance Specialist				**	**	**
75B	Personnel Administration Specialist	-.35	-.25	**	**	**	**

* Because attrition is a dichotomous variable, the resulting biserial correlation coefficients cannot be translated to Pearson product-moment coefficients. Thus, correlations should underestimate the relationship between attrition and AFQT scores.

** Sample size too small (N<30).

Table 31

Uncorrected Correlations Between Aptitude Composite Scores and Measures of Training Performance Across Racial/Ethnic Groups for Four MOSs

MOS		Aptitude Composite	Final Course Grade			Attrition*		
			White	Black	Hispanic	White	Black	Hispanic
05C	Radio Teletypewriter Operator	SC	.15	.27	**	.13	.12	**
31M	Multichannel Communications Operator	EL	.55	.21	**	.17	.32	**
73C	Finance Specialist	CL	.45	.34	**			
75B	Personnel Administration Specialist	CL	.25	.20	**			

MOS		Aptitude Composite	Time-to-Complete			Peer Nomination		
			White	Black	Hispanic	White	Black	Hispanic
05C	Radio Teletypewriter Operator	SC	-.11	-.22	**			
31M	Multichannel Communications Operator	EL				.16	.36	**
73C	Finance Specialist	CL				**	**	**
75B	Personnel Administration Specialist	CL	-.36	-.38	**	**	**	**

* Because attrition is a dichotomous variable, the resulting biserial correlation coefficients cannot be translated to Pearson product-moment coefficients. Thus correlations should underestimate the relationship between attrition and aptitude composite scores.

** Sample size too small (N<30).

Appendix J

SERVICE ENLISTMENT STANDARDS

Table 32. Service Enlistment Standards by Sex, Test Form, and Level of Education

Table 32

Service Enlistment Standards by Sex, Test Form, and Level of Education

Males

Service/Education	ASVAB 6, 7 Score	ASVAB 8,9,10 Directed Minimum Score*	ASVAB 8,9,10 Operational Score**
Army			
High School Diploma Graduate			
AFQT	16	12	16
Aptitude Composite	1-90	1-80	1-85
Non-High School Graduate			
AFQT	31	17	31
Aptitude Composite	2-90s	2-80s	2-85s
Marine Corps			
High School Diploma Graduate			
AFQT	21	14	21
Aptitude Composite	GT-80	GT-69	GT-80
Non-High School Graduate			
AFQT	21	14	21
Aptitude Composite	GT-95	GT-85	GT-95

Females

Service/Education	ASVAB 6, 7 Score	ASVAB 8,9,10 Directed Minimum Score *	ASVAB 8,9,10 Operational Score **
Army			
Marine Corps		Same as for Males	
High School Diploma Graduate			
AFQT	50	41	50
Non-High School Graduate			
AFQT	Not Eligible	Not Eligible	Not Eligible

* OASD(MRA&L) directed Services to adjust enlistment standards under ASVAB 8/9/10 to qualify same types of people who have previously qualified for enlistment under ASVAB 6/7. Scores listed below are equivalent to scores under ASVAB 6/7.

** Operational score currently being used by recruiters is based upon Service estimates of recruiting market. Army and Marine Corps state that they have established minimum enlistment standards to conform with DoD directive.

Appendix K

ASSESSMENT OF JOB PERFORMANCE MEASURES -- SYSTEM REQUIREMENTS

A major purpose of the military personnel system is to facilitate the acquisition and utilization of personnel in ways that enhance the effectiveness and efficiency of the military force. Within the system, various tests are employed for selection, classification, training, assignment, promotion, and retention. Measures used for selection and classification should predict training performance, which in turn should be related to job performance. In order to allow for the validation of predictive measures, training and job performance measures, both at the individual and unit level, must not only lead to accurate decisions regarding competence (i.e., acceptable versus unacceptable performance), but also must display sufficient variance for demonstrating a strong relationship (correlation) between selection and classification measures and job performance.

A test which is to be used to make decisions regarding competence would be most appropriate for making the kind of Go/No Go decisions required in military training and for determining combat readiness. Since such a test would have to reliably discriminate among levels of performance at or near the pass point, but not among various higher (or lower) levels of performance, these tests may be brief and may have relatively small variance. Alternatively, a test to be used as a criterion for validating selection and classification measures must provide continuous measures of levels of competence along a wider performance scale. Such tests would tend to be longer and have more variance, providing better estimates of correlations with predictor measures.

Organizational Considerations

Organizational constraints within a military setting must be considered in evaluating measures of training and job performance. Necessarily, factors such as mission, doctrine, and equipment must be considered in specifying appropriate measures. Specification of criteria is complicated by the dynamic nature of these factors: missions change, doctrine is modified, and new equipment is introduced into the defense inventory. Thus, performance measures must take into account both present and future requirements.

Because most enlistees serve in the military for a period of three years and since training is an ongoing process, consideration must be given as to when the optimal time is to assess performance. Thus, length of service at time of test may influence performance (Vineberg & Taylor, 1972). Numerous studies have shown that performance early in the learning period does not necessarily correlate highly with later performance (Kornhauser, 1923; Blankenship and Taylor, 1938; McGehee, 1948; Smith and Gold, 1956; Ghiselli and Haire, 1960; Bass, 1962; Prien, 1966; MacKinney, 1967). Therefore, performance assessments should be made only on individuals who have been in their current job for some minimum length of time. Furthermore, if supervisory ratings are to be used as criteria of performance, it is critical to insure that supervisors are familiar with personnel. These constraints

reduce the utility of job performance measures when they are routinely applied on a fixed schedule without regard to personnel or equipment changes.

Another factor complicating job performance measurement is the adoption of new equipment into the defense inventory, often making some tasks obsolete and introducing others. Less obviously, such innovations may modify the performance requirements of some tasks. New operational procedures may necessitate changes in task accuracy or speed. The tolerance of adjustment, calibration, or alignment procedures may change as well. The dynamic nature of jobs, therefore, requires that measures of job performance be capable of adapting to such changes in the job in a timely manner.

The sheer number of military occupational specialties for which job performance measures must be developed causes severe logistical problems for the Army. To compound this problem, not everyone within an MOS does all the tasks prescribed for that MOS. A great deal of job specialization takes place depending on the unit of assignment. The issue here is whether job performance measurement should focus on those tasks which an individual soldier typically performs or upon all the tasks the soldier might be called upon to perform.

Finally, because of the wide geographic dispersion of soldiers in various MOSs, performance measurement instruments must be transportable and must be flexible enough to be administered in a variety of circumstances. Further, the obtained information must be managed in a way that permits tracking of individuals on a longitudinal basis.

Cost Considerations

In selecting a particular performance measure from the various types available, it is important to compare their cost effectiveness. Significant costs may be incurred during test development, administration, and data processing. Efficient data collection with minimal disruption of normal duties is desired. In addition, since test results are to be used for decision making, the performance measurement system should permit timely processing of the required information to the required levels (e.g., when instruments are administered with the intent of evaluating the individual soldier, evaluating the training system, or obtaining an estimate of individual competences within operational units).

Technical Considerations

In evaluating various types of job performance measures, major technical issues include validity, reliability, equity and power to discriminate among adequate and inadequate performers. Validity is the first requirement for a performance measure and one which involves the elucidation of criteria for job performance. However, except under condition of war, the ultimate criteria of the effectiveness of the armed forces cannot be observed. In peacetime, assessment of combat effectiveness must rely on substitute measures which reflect combat readiness. The validity of such measures depends upon their fidelity to the requirements and conditions of combat and on the extent to which they measure all aspects of job performance. Command post exercises, field maneuvers, Army Readiness Training Exercises (ARTEP), and Skill Qualification Tests are forms of simulation aimed at assessment of combat readiness. To the extent that such simulations represent

(simulate) critical components of combat effectiveness, they are content valid and can serve as effective estimators of combat effectiveness. Selection and training measures that predict success in these simulations can then be considered both valid and useful.

Job performance measures, in order to demonstrate validity, must be reliable. A measure's reliability may be assessed in terms of agreement between different evaluations, or with different but similar measures. Unfortunately, there is frequently an inverse relationship between fidelity and reliability (Fitzpatrick and Morrison, 1969). Often, the more performance tests resemble actual job conditions the less control the examiner has over the situation. In such cases different soldiers are likely to be required to perform different tasks or perform the same tasks under different conditions, and thus there will be more variability in potential responses and less justification for comparing scores.

User acceptability is another important characteristic of performance measures which must receive consideration. Obviously, regardless of the predictive power of the particular measures, they must be acceptable to the user or they will be rejected. Key factors for user acceptance are "face" validity, administrative convenience, and associated costs of time, equipment, and personnel required for administration.

Finally, evidence must be presented to demonstrate the equity or fairness of performance measures. For example, measurement instruments should be free from biases which favor one group at the expense of another on some irrelevant basis (e.g., written tests which require reading levels greater than those needed on the job). Equity must also be maintained with regard to ethnic composition, sex, and age of recruits.

Comparison of Various Performance Measures

The major types of job performance measures -- direct performance tests, simulations, written tests, ratings -- have different strengths and weaknesses with regard to the technical and cost considerations discussed above. Advantages of direct performance testing lie in its high fidelity and usefulness as a learning tool. Feedback can be given to test takers or used to develop individual programs of instruction in weak areas. Typically, performance testing is comprehensive for those tasks which are evaluated but, because of cost, only a few of the tasks in the job domain are tested. Reliability may be a problem also because under actual job conditions the test environment is typically not completely controlled. Simulations sacrifice some of the fidelity of actual performance tests in return for a greater degree of control and reliability. They are usually more expensive to control and administer, however.

Job knowledge tests, in contrast, frequently cover a larger portion of the job domain than do performance tests but with less fidelity. While such written tests can have significant development costs they are still much less expensive overall than performance tests. They are much cheaper to administer and score, and there is much less subjectivity in the scoring process, resulting in greater reliability. More sophisticated statistical analyses can be performed on data from written tests and it is easier to select items based on their ability to discriminate between good and poor performers.

Ratings are the most subjective of the performance measures. Although some types (i.e., behaviorally anchored scales) may be rather expensive to develop, ratings are generally the least expensive method of performance appraisal. The trade-off is, however, decreased reliability and validity.

REFERENCES

Bass, B.M. Further evidence on the dynamic character of criteria. Personnel Psychology, 1962, 15, 93-97.

Blankenship, A.B., & Taylor, H.R. Prediction of vocational proficiency in three machine operations. Journal of Applied Psychology, 1938, 22, 518-526.

Fitzpatrick, R. & Morrison, E.J. Performance and product evaluation. In Thorndike, R.L. (Ed.), Educational Measurement (2nd Ed.), Washington, D.C.: American Council on Education, 1971.

Ghiselli, E.E., & Haire, M. The validation of selection tests in the light of the dynamic character of criteria. Personnel Psychology, 1960, 13, 225-231.

Kornhauser, A.W. A statistical study of a group of specialized office workers. Journal of Personnel Research, 1923, 2, 103-123.

MacKinney, A.C. An assessment of performance change: An inductive example. Organizational Behavior and Human Performance, 1967, 3, 56-72.

McGehee, W. Cutting training waste. Personnel Psychology, 1948, 1, 331-340.

Prien, E.P. Dynamic character of criteria: Organizational change. Journal of Applied Psychology, 1966, 50, 501-504.

Smith, P.C., & Gold, R.A. Prediction of success from examination of performance during the training period. Journal of Applied Psychology, 1956, 40, 83-86.

Vineberg, R., & Taylor, E.N. Performance in four Army jobs by men at different aptitude (AFQT) levels: 3. the relationship of AFQT and job experience to job performance (HumRRO-TR-72-22). Alexandria, Virginia: Human Resources Research Organization, August 1972.

Appendix L

EXPERIMENTAL JOB PERFORMANCE MEASUREMENT INSTRUMENTS

Appendix L consists of copies of the first two pages of the three experimental job performance measurement instruments for MOSs 11B, 31M, and 75B which were developed for this study.* They are as follows:

- Occupational Survey Form
- Task Difficulty Rating Form
- Job Performance Rating Form

---

* For a complete list of tasks tested by these three instruments, the reader should refer to the Soldier's Manuals for the specific MOS of interest.

OCCUPATIONAL SURVEY FOR 11B

SOLDIER NAME _____ SSN _____

WHEN DID YOU COMPLETE OSUT/AIT TRAINING FOR MOS? _____

GRADE _____ TOE/TDA DUTY POSITION (JOB TITLE) _____

HOW LONG HAVE YOU BEEN WORKING IN YOUR PRESENT JOB? _____ MONTHS

HAVE YOU TAKEN AN SQT? _____ IF SO, WHEN? _____/_____ WHERE? _____
MONTH YEAR

The 11B Soldier's Manual lists, for each skill level, the tasks that each soldier is expected to perform on the job. In fact, you probably do not do all of these tasks in your present job. You may not have the equipment or you may specialize on the job.

- o On the next few pages all skill level 1 tasks except common soldiering tasks are listed.
- o Task statements are grouped under functional area titles and under a few duty positions which have unique tasks.
- o Read through this list and place a check next to only those tasks which you perform <u>now</u> in your present job (duty position).

Your responses will <u>not</u> be used to evaluate you or your unit.

COMBAT TECHNIQUES BASIC INDIVIDUAL TECHNIQUES	Check (√) only tasks you perform now in your job
Move as a member of a fire team	
Move under direct fire	
React to indirect fire	
React to flares	
Move over, through, or around obstacles	
Estimate range	
Select temporary battlefield positions	
Construct individual fighting position	
Use visual signals to control movement (dismounted)	
CAMOUFLAGE, COVER, AND CONCEALMENT Camouflage/conceal self and individual equipment	
Camouflage/conceal equipment	
Camouflage/conceal defensive positions	
Clear fields of fire	

TASK DIFFICULTY RATING FORM FOR 11B

SUPERVISOR NAME _____ GRADE _____

HOW LONG HAVE YOU BEEN IN YOUR PRESENT JOB? _____ MONTHS

The 11B Soldier's Manual lists, for each skill level, the tasks that each soldier is expected to perform on the job. Some of these tasks are more difficult to perform than others. When we say a task is difficult we mean that it takes an average skill level 1 soldier a long time to learn to perform well.

- On the next few pages all skill level 1 tasks, except common soldiering tasks, are listed.
- Task statements are grouped under functional area titles and under a few duty positions which have unique tasks.
- Rate the difficulty of each task by circling the appropriate number to the right of each task statement. Remember, a difficult task is one that takes skill level 1 soldiers a long time to learn to perform well.

Page 1	Task Difficulty Rating for 11B				
Task	Task takes short time to learn to perform		Task takes moderate time to learn to perform		Task takes long time to learn to perform
COMBAT TECHNIQUES					
BASIC INDIVIDUAL TECHNIQUES					
Move as a member of a fire team	1	2	3	4	5
Move under direct fire	1	2	3	4	5
React to indirect fire	1	2	3	4	5
React to flares	1	2	3	4	5
Move over, through, or around obstacles	1	2	3	4	5
Estimate range	1	2	3	4	5
Select temporary battlefield positions	1	2	3	4	5
Construct individual fighting position	1	2	3	4	5
Use visual signals to control movement (dismounted)	1	2	3	4	5
CAMOUFLAGE, COVER, AND CONCEALMENT					
Camouflage/conceal self and individual equipment	1	2	3	4	5
Camouflage/conceal equipment	1	2	3	4	5
Camouflage/conceal defensive positions	1	2	3	4	5
Clear fields of fire	1	2	3	4	5

JOB PERFORMANCE RATING FORM FOR 11B

SUPERVISOR NAME _____ GRADE _____ DUTY MOS _____

SOLDIER's NAME _____ GRADE _____ DUTY MOS _____

SOLDIER'S TOE/TDA DUTY POSITION (JOB TITLE) _____

HOW LONG HAVE YOU BEEN SUPERVISING THIS SOLDIER? _____ MONTHS

The 11B Soldier's Manual lists, for each skill level, the tasks that each soldier is expected to perform on the job. In fact, you have probably not observed the soldier performing all of these tasks.

- On the next few pages, skill level 1 tasks are grouped under functional area titles and by duty position.
- In his present job, the soldier probably does not perform all of the tasks listed. You may not have the equipment in your unit or the soldier may specialize on the job. Place a check next to only those tasks which the soldier performs now in his present job.
- Rate the soldier's performance on each task which you have ever observed the soldier perform, or observed the results of the soldier's performance, by circling the appropriate number to the right of each task statement. The tasks on which you are rating the soldier may or may not be the same ones you checked before.

Page 1 Soldier's performance on tasks which you have observed soldier perform:						
Task	Tasks which soldier performs now in his present job	is below required standard		meets required standard		exceeds required standard
COMBAT TECHNIQUES						
BASIC INDIVIDUAL TECHNIQUES						
Move as a member of a fire team		1	2	3	4	5
Move under direct fire		1	2	3	4	5
React to indirect fire		1	2	3	4	5
React to flares		1	2	3	4	5
Move over, through, or around obstacles		1	2	3	4	5
Estimate range		1	2	3	4	5
Select temporary battlefield positions		1	2	3	4	5
Construct individual fighting position		1	2	3	4	5
Use visual signals to control movement (dismounted)		1	2	3	4	5
CAMOUFLAGE, COVER, AND CONCEALMENT						
Camouflage/conceal self and individual equipment		1	2	3	4	5
Camouflage/conceal equipment		1	2	3	4	5
Camouflage/conceal defensive positions		1	2	3	4	5
Clear fields of fire		1	2	3	4	5

OCCUPATIONAL SURVEY FOR 31M

SOLDIER NAME _____ SSN _____

WHEN DID YOU COMPLETE OSUT/AIT TRAINING FOR YOUR MOS? _____

GRADE _____ TOE/TDA DUTY POSITION (JOB TITLE) _____

HOW LONG HAVE YOU BEEN WORKING IN YOUR PRESENT JOB? _____ MONTHS

HAVE YOU TAKEN AN SQT? _____ IF SO, WHEN? _____/_____ WHERE? _____
MONTH YEAR

The 31M Soldier's Manual lists, for each skill level, the tasks that each soldier is expected to perform on the job. In fact, you probably do not do all of these tasks in your present job. You may not have the equipment or you may specialize on the job.

- o On the next few pages all skill level 1 tasks except common soldiering tasks are listed.
- o Task statements are grouped under equipment area titles.
- o Read through this list and place a check next to only those tasks which you perform <u>now</u> in your present job (duty position).

Your responses will <u>not</u> be used to evaluate you or your unit.

Page 1 LOW CAPACITY EQUIPMENT	Check (√) only tasks you perform now in your job
Install Generator Set, 3 KW, PU-625/G or PU-628/G	
Operate Generator Set, 3 KW, PU-625/G or PU-628/G	
Perform Operator's Daily Preventive Maintenance on Generator Set, 3 KW, PU-625/G or PU-628/G	
Troubleshoot Generator Set, 3 KW, PU-625/G or PU-628/G	
Install Radio Terminal Set, AN/TRC-145(V)	
Operate Radio Terminal Set, AN/TRC-145(V)	
Perform Operator's Daily Preventive Maintenance on Radio Terminal Set, AN/TRC-145(V)	
Troubleshoot Radio Terminal Set, AN/TRC-145(V)	

TASK DIFFICULTY RATING FORM FOR 31M

SUPERVISOR NAME _____ GRADE _____

HOW LONG HAVE YOU BEEN WORKING IN YOUR PRESENT JOB? _____ MONTHS

The 31M Soldier's Manual lists, for each skill level, the tasks that each soldier is expected to perform on the job. Some of these tasks are more difficult to perform than others. When we say a task is difficult we mean that it takes a long time for an average skill level 1 soldier to learn to perform well.

- On the next few pages, all skill level 1 tasks are listed.
- Task statements are grouped under equipment area titles.
- Rate the difficulty of each task by circling the appropriate number to the right of each task statement. Remember, a difficult task is one that takes skill.
- Finally, each time a piece of equipment is mentioned for the first time, please place a check in the appropriate box if you have that equipment in your unit.

Page 1	Task Difficulty Rating for 31M					
Task	Check (√) if you have equipment	Task takes short time to learn to perform		Task takes moderate time to learn to perform		Task takes long time to learn to perform
LOW CAPACITY EQUIPMENT Install Generator Set, 3 KW, PU-625/G or PU-628/G		1	2	3	4	5
Operate Generator Set, 3 KW, PU-625/G or PU-628/G		1	2	3	4	5
Perform Operator's Daily Preventive Maintenance on Generator Set, 3 KW, PU-625/G or PU-628/G		1	2	3	4	5
Troubleshoot Generator Set, 3 KW, PU-625/G or PU-628/G		1	2	3	4	5
Install Radio Terminal Set, AN/TRC-145(V)		1	2	3	4	5
Operate Radio Terminal Set, AN/TRC-145(V)		1	2	3	4	5
Perform Operator's Daily Preventive Maintenance on Radio Terminal Set, AN/TRC-145(V)		1	2	3	4	5
Troubleshoot Radio Terminal Set, AN/TRC-145(V)		1	2	3	4	5

JOB PERFORMANCE RATING FORM FOR 31M

SUPERVISOR NAME _____ GRADE _____ DUTY MOS _____

SOLDIER'S NAME _____ GRADE _____ DUTY MOS _____

SOLDIER'S TOE/TDA DUTY POSITION (JOB TITLE) _____

HOW LONG HAVE YOU BEEN SUPERVISING THIS SOLDIER? _____ MONTHS

The 31M Soldier's Manual lists, for each skill level, the tasks that each soldier is expected to perform on the job. In fact, you have probably not observed the soldier performing all of these tasks.

- On the next few pages, skill level 1 tasks are grouped under equipment area titles.
- In his present job, the soldier probably does not perform all of the tasks listed. You may not have the equipment in your unit or the soldier may specialize on the job. Place a check next to only those tasks which the soldier performs now in his present job.
- Rate the soldier's performance on each task which you have ever observed the soldier perform, or observed the results of the soldier's performance by circling the appropriate number to the right of each task statement. The tasks on which you are rating the soldier may or may not be the same ones you checked before.

Page 1 Task	Tasks which soldier performs now in his present job	Soldier's performance of task:				
		is below required standard		meets required standard		exceeds required standard
LOW CAPACITY EQUIPMENT Install Generator Set, 3 KW, PU-625/G or PU-628/G		1	2	3	4	5
Operate Generator Set, 3 KW, PU-625/G or PU-628/G		1	2	3	4	5
Perform Operator's Daily Preventive Maintenance on Generator Set, 3 KW, PU-625/G or PU-628/G		1	2	3	4	5
Troubleshoot Generator Set, 3 KW, PU-625/G or PU-628/G		1	2	3	4	5
Install Radio Terminal Set, AN/TRC-145(V)		1	2	3	4	5
Operate Radio Terminal Set, AN/TRC-145(V)		1	2	3	4	5
Perform Operator's Daily Preventive Maintenance on Radio Terminal Set, AN/TRC-145(V)		1	2	3	4	5
Troubleshoot Radio Terminal Set, AN/TRC-145(V)		1	2	3	4	5

OCCUPATIONAL SURVEY FOR 75B

SOLDIER NAME _____ SSN _____

WHEN DID YOU COMPLETE OSUT/AIT TRAINING FOR YOUR MOS? _____

GRADE _____ TOE/TDA DUTY POSITION (JOB TITLE) _____

HOW LONG HAVE YOU BEEN WORKING IN YOUR PRESENT JOB? _____ MONTHS

HAVE YOU TAKEN AN SQT? _____ IF SO, WHEN? ____/____ WHERE? _____
MONTH YEAR

The 75B Soldier's Manual lists, for each skill level, the tasks that a soldier is expected to perform on the job. In fact, you probably do not do all of these tasks in your present job. You may specialize on the job or your unit may not be responsible for all tasks listed.

- o On the next few pages all skill level 1 tasks except common soldiering tasks are listed.
- o Task statements are grouped under functional area titles.
- o Read through this list and place a check next to only those tasks which you perform <u>now</u> in your present job.

Your responses will <u>not</u> be used to evaluate you or your unit.

Task	Check (✓) only tasks *you perform now* in your job
SIDPERS INPUT Prepare SIDPERS Change Reports, DA Form 3728	
Prepare Organizational Strength Reports, DA Form 3732	
Determine reportable changes to SIDPERS	
Determine the category of SIDPERS changes	
Determine reportable duty status codes	
Determine required documentation for SIDPERS transactions	
Prepare a Request for a Personnel Action, DA Form 4187, Parts I, II, IV, V	
SIDPERS OUTPUT REPORTS Process the Unit Manning Report (UMR)	
Process the Personnel Transaction Register by Originator (PTRO)	
Process the Unit Personnel Accountability Notice (UPAN)	

TASK DIFFICULTY RATING FORM FOR 75B

SUPERVISOR NAME _____ GRADE _____

HOW LONG HAVE YOU BEEN WORKING IN YOUR PRESENT JOB? _____ MONTHS

The 75B Soldier's Manual lists, for each skill level, the tasks that each soldier is expected to perform on the job. Some of these tasks are more difficult to perform than others. When we say a task is difficult we mean that it takes a long time for an average skill level 1 soldier to learn to perform well.

- o On the next few pages, all skill level 1 tasks are listed.
- o Task statements are grouped under functional area titles.
- o Rate the difficulty of each task by circling the appropriate number to the right of each task statement. Remember, a difficult task is one that takes skill level 1 soldiers a long time to learn to perform well.

Page 1	Task Difficulty Rating for 75B				
Task	Task takes short time to learn to perform		Task takes moderate time to learn to perform		Task takes long time to learn to perform
SIDPERS INPUT					
Prepare SIDPERS Change Reports, DA Form 3728	1	2	3	4	5
Prepare Organizational Strength Reports, DA Form 3732	1	2	3	4	5
Determine reportable changes to SIDPERS	1	2	3	4	5
Determine the category of SIDPERS changes	1	2	3	4	5
Determine reportable duty status codes	1	2	3	4	5
Determine required documentation for SIDPERS transactions	1	2	3	4	5
Prepare a Request for a Personnel Action, DA Form 4187, Parts I, II, IV, V	1	2	3	4	5
SIDPERS OUTPUT REPORTS					
Process the Unit Manning Report (UMR)	1	2	3	4	5
Process the Personnel Transaction Register by Originator (PTRO)	1	2	3	4	5
Process the Unit Personnel Accountability Notice (UPAN)	1	2	3	4	5

JOB PERFORMANCE RATING FORM FOR 75B

SUPERVISOR NAME _____ GRADE _____ DUTY MOS _____

SOLDIER'S NAME _____ GRADE _____ DUTY MOS _____

SOLDIER'S TOE/TDA DUTY POSITION (JOB TITLE) _____

HOW LONG HAVE YOU BEEN SUPERVISING THIS SOLDIER? _____ MONTHS

The 75B Soldier's Manual lists, for each skill level, the tasks that each soldier is expected to perform on the job. In fact, you have probably not observed the soldier perform on the job. In fact, you have probably not observed the soldier performing all of these tasks.

- On the next few pages, skill level 1 taks are grouped under functional area titles.
- In his present job, the soldier probably does not perform all of the tasks listed. The soldier may specialize on the job. Place a check next to only those tasks which the soldier performs now in his present job.
- Rate the soldier's performance on each task which you have ever observed the soldier perform, or observed the results of the soldier's performance, by circling the appropriate number to the right of each task statement. The tasks on which you are rating the soldier may or may not be the same ones you checked before.

Page 1	Soldier's performance on tasks which you have observed soldier perform:					
Task	Tasks which soldier performs now in his present job	is below required standard		meets required standard		exceeds required standard
SIDPERS INPUT						
Prepare SIDPERS Change Reports, DA Form 3728		1	2	3	4	5
Prepare Organizational Strength Reports, DA Form 3732		1	2	3	4	5
Determine reportable changes to SIDPERS		1	2	3	4	5
Determine the category of SIDPERS changes		1	2	3	4	5
Determine reportable duty status codes		1	2	3	4	5
Determine required documentation for SIDPERS transactions		1	2	3	4	5
Prepare a Request for a Personnel Action, DA Form 4187, Parts I, II, IV, V		1	2	3	4	5
SIDPERS OUTPUT REPORTS						
Process the Unit Manning Report (UMR)		1	2	3	4	5
Process the Personnel Transaction Register by Originator (PTRO)		1	2	3	4	5
Process the Unit Personnel Accountability Notice (UPAN)		1	2	3	4	5